E2 01 573

2

AD-A269 104

PL-TR-93-2100

# SERCAA CLOUD ANALYSIS INTEGRATION: DESIGN CONCEPTS AND INTERACTION WITH CLOUD FORECAST MODELS

**Thomas M. Hamill**
**Ross Hoffman**

**Atmospheric and Environmental Research, Inc.**
**840 Memorial Drive**
**Cambridge, MA 02139**

**16 April 1993**

**Scientific Report No. 1**



**Approved for public release; distribution unlimited**

**PHILLIPS LABORATORY**
**Directorate of Geophysics**
**AIR FORCE MATERIEL COMMAND**
**HANSCOM AIR FORCE BASE, MA 01731-3010**

93-18035

"This technical report has been reviewed and is approved for publication."

ALLAN J. BUSSEY
Contract Manager

JOSEPH W. SNOW
Chief, Satellite Meteorology Branch
Atmospheric Sciences Division

ROBERT A. McCLATCHEY
Director, Atmospheric Sciences Division

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify PL/TSI, 29 Randolph Street, Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

| REPORT DOCUMENTATION PAGE | *Form Approved OMB No. 0704-0188* |
|---|---|



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 16 April 1993 | Scientific No. 1 |

| 4. TITLE AND SUBTITLE | 5. FUNDING NUMBERS |
|---|---|
| SERCAA Cloud Analysis Integration: Design Concepts and Interaction with Cloud Forecast Models | Contract F19628-92-C-0149<br>PE 35160F<br>PR DSPO<br>TA GR<br>WU AA |
| 6. AUTHOR(S)<br>Thomas M. Hamill and Ross Hoffman | |
| 7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)<br>Atmospheric and Environmental Research, Inc.<br>840 Memorial Drive<br>Cambridge, MA 02139 | 8. PERFORMING ORGANIZATION REPORT NUMBERS<br>(none) |
| 9. SPONSORING / MONITORING AGENCY NAMES(S) AND ADDRESS(ES)<br>Phillips Laboratory<br>29 Randolph Road<br>Hanscom AFB, MA 01731-3010<br>Contract Manager: Allan Bussey/GPAS | 10. SPONSORING / MONITORING AGENCY REPORT NUMBER<br>PL-TR-93-2100 |

11. SUPPLEMENTARY NOTES

| 12a. DISTRIBUTION / AVAILABILITY STATEMENT | 12b. DISTRIBUTION CODE |
|---|---|
| Approved for public release; distribution unlimited | |

13 ABSTRACT *(Maximum 200 words)*

AER, in conjunction with the Phillips Lab, is prototyping a new cloud analysis model to eventually replace the Real-Time Nephanalysis operating at the Air Force Global Weather Central. This new cloud analysis model will process multisource satellite data, geostationary and polar-orbiting, civilian and military. The output will be used primarily to initialize cloud forecast models but may be used also for climate research and other applications. This report will review the customer requirements for cloud analysis data, the existing and future cloud forecast algorithms, and how the new cloud analysis model can integrate the multisource satellite data into one coherent analysis useful to all users. The likely approach for this "analysis integration" will be to blend optimum interpolation techniques, common in numerical weather analysis, with a knowledge base commonly used in artificial intelligence.

| 14. SUBJECT TERMS | | | 15. NUMBER OF PAGES |
|---|---|---|---|
| artificial intelligence | cloud analysis | cloud forecasting | 60 |
| | satellite meteorology | optimum interpolation | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | SAR |

NSN 7540-01-280-5500

Standard Form 298 (rev 2-89)
Prescribed by ANSI Std. Z39-18
298-102

## Table of Contents

Table of Contents
(continued)

## List of Figures

## List of Tables

## 1. Introduction

By October 1997 the Air Force Global Weather Central (AFGWC) plans to replace much of its existing computer network, and with it, implement a new suite of global cloud analysis and forecast software (AFGWC, 1992). The analysis and forecast output is used by a variety of high-priority customers for wartime and peacetime operations and planning. In addition, the analysis database is permanently archived for climatological studies, weapons design, technique development, and other research efforts. However, AFGWC's existing cloud analysis and cloud forecast algorithms are out-of-date (Hamill et al., 1992). The current cloud analysis model, called the Real-Time Nephanalysis, or RTNEPH, can process only VIS and IR satellite data, whereas current military satellites also include a suite of microwave imagers, and civilian satellites include 5-channel, multispectral imagery (from the AVHRR radiometer aboard the polar-orbiting TIROS series) and high-temporal resolution data (the geostationary satellites). The forecast algorithms are similarly out-of-date, partly limited by the sub-optimum quality of the RTNEPH and partly by the low-resolution, oversimplified numerical forecast algorithms.

Two years ago the U.S. Air Force Phillips Laboratory's Satellite Meteorology Branch (PL/GPAS) submitted a proposal to improve the RTNEPH algorithms under the Strategic Environmental Research and Development Program (SERDP). The PL/GPAS project was given the name Support of Environmental Requirements for Cloud Analysis and Archive (SERCAA). The SERDP program, created by Congress, funds R&D efforts to transfer Department of Defense technology to help with environmental issues such as global warming. There is a cloud analysis database available to the civilian climate community, called ISCCP, or the International Satellite Cloud Climatology Project (Rossow et al., 1985; Schiffer and Rossow, 1985). However, this database is very coarse in both temporal and spatial resolution, and does not have many of the cloud parameters oft-desired by users, such as cloud-base height. Thus, PL/GPAS proposed a 3-year research effort to update the old RTNEPH technology to use new data sources and processing algorithms. This new technology could be used by the Air Force or incorporated into a new civilian operational nephanalysis. Hereafter we will call the new model the MSNEPH, or multispectral/multisource nephanalysis.

Because of the system deficiencies at AFGWC noted above, the Air Force was particularly interested in the SERCAA project. The Air Force worked with SERDP program managers to ensure funding for the project and to fold the SERCAA work into the planned upgrade of AFGWC's computers. Since the new hardware and software must be operational by 1997, this requires the SERCAA project, which began in November 1992, to deliver the crucial design enhancements very early.

This quick turnaround requires that AFGWC and PL/GPAS work closely to mesh the new SERCAA algorithm design with the existing and the planned cloud forecast models. Because the existing cloud forecast models are also antiquated, simply retrofitting the new nephanalysis algorithms to the RTNEPH database structure may not be desirable; there may be additional information useful to the forecast model, or conversely, information stored in the current RTNEPH that does not improve the cloud forecasts. At the same time, the SERCAA design and the Air Force should account for other potential users who may desire a significantly different database. The climate users, weapons designers, and tactical military users may have different priorities from AFGWC's main customer, which is interested in the nephanalysis mostly as a model for initializing cloud forecast models. Piecing together a new nephanalysis to satisfy all these needs will be a complex task.

This technical report will review some of the important design issues and tradeoffs for the MSNEPH. It will concentrate on the links with AFGWC's current and future forecast models and the design of the module which will blend together the often disparate satellite cloud analyses - what we call "analysis integration." This document is meant to stimulate up-front discussion among designers and users to ensure the best system design possible. This technical report will review the existing AFGWC cloud analysis and forecast cycle and its deficiencies (Section 2) and review the various customer requirements as they are currently known (Section 3). Since most of the basic satellite analysis technologies are well-known and documented, these will not be described here. Rather in this document we will review the possible cloud forecast technologies (Section 4), and how to design the new cloud analysis integration algorithm so it provides both a robust standalone data base and link to the forecast models (Section 5). A range of integration algorithms will be proposed and examined for advantages and disadvantages, and one compromise design will be proposed that we believe will best meet the various requirements. Conclusions will be provided in Section 6.

## 2. Current AFGWC Cloud Forecast Cycle

The cloud forecast model process at AFGWC has grown by amalgamation. As a result, the system is somewhat odd. Further, it is not thoroughly documented. Figure 1 diagrams the main processes and databases in producing AFGWC's cloud forecasts. The top part, the ingest of the raw DMSP (Defense Meteorological Satellite Processor) data, its cartesian rectification, and processing into a cloud analysis by the RTNEPH was recently described in detail in Hamill et al. (1992). For the rest of the process, a basic technical report is available (Crum, 1987), though much has changed since its publication. There are currently three main forecast models, the 5-Layer model, the High-Resolution Cloud Prog (HRCP), and TRONEW, a diurnal persistence model. SAVDOX software synthesizes cloud forecasts. Below, the basics of the existing cloud forecast cycle are documented and the known deficiencies described.

### *2.1 Description of Forecast Models and Synthesis Software*

#### *2.1.1 5-Layer*

The 5-Layer trajectory model produces cloud forecasts from 3 to 48 hours for the midlatitudes. The model has 5 vertical layers, one at the top of the boundary layer, called the gradient layer, and four in the free atmosphere at 850, 700, 500, and 300 mb. Its horizontal resolution is half-mesh, or 190.5 km (at 60 degrees latitude). The 5-Layer model is only run over an area called the GWC octagon, a stop-sign shaped domain covering the mid- and high-latitudes (Figure 2 - N. Hem). The purpose of the 5-Layer is to provide mid-latitude cloud forecasts for forecasts over 9 hours in length. The forecast technique is rather simple; the RTNEPH cloud information is first compacted

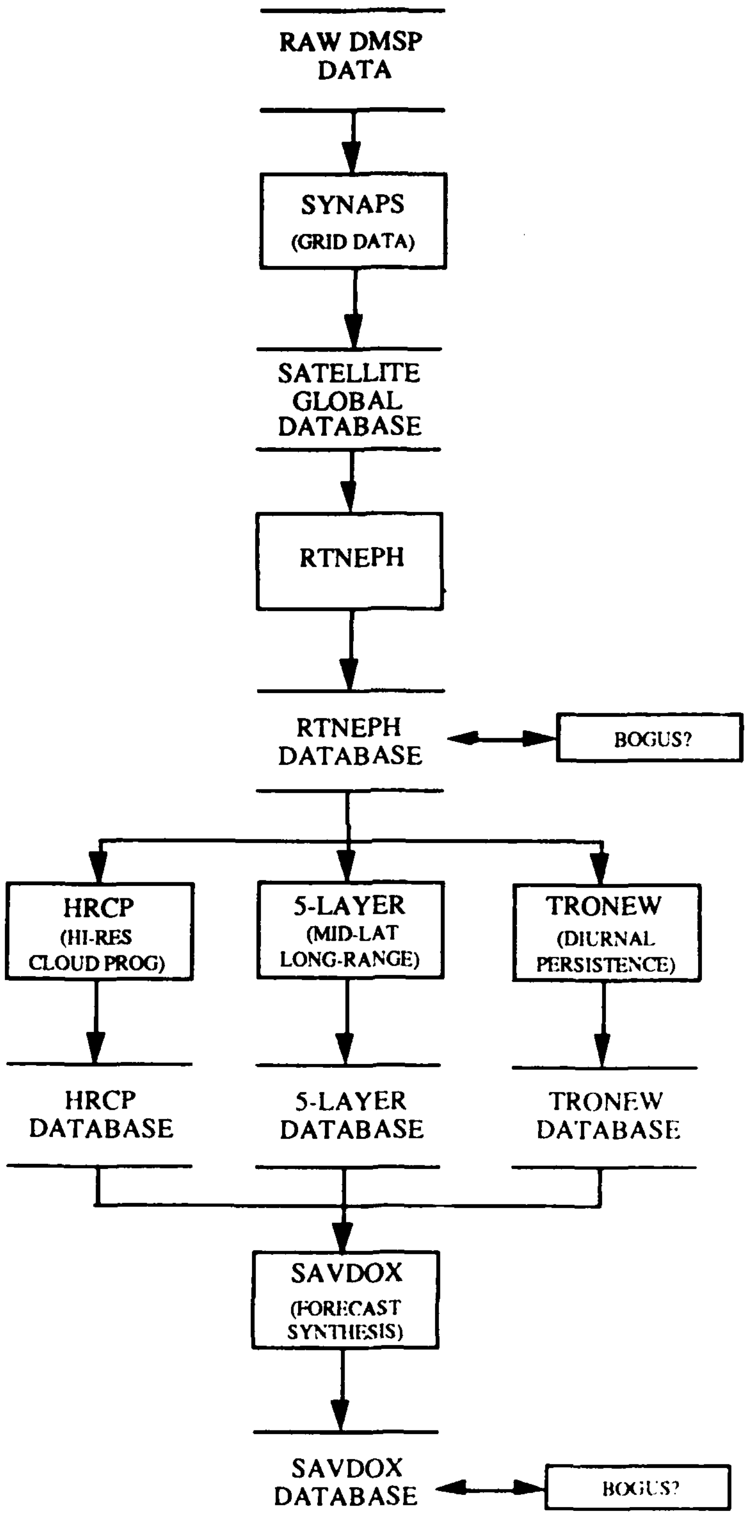


Figure 1: Data and processing flow for the AFGWC cloud forecast cycle. Boxes represent algorithms, and parallel lines the data stores.

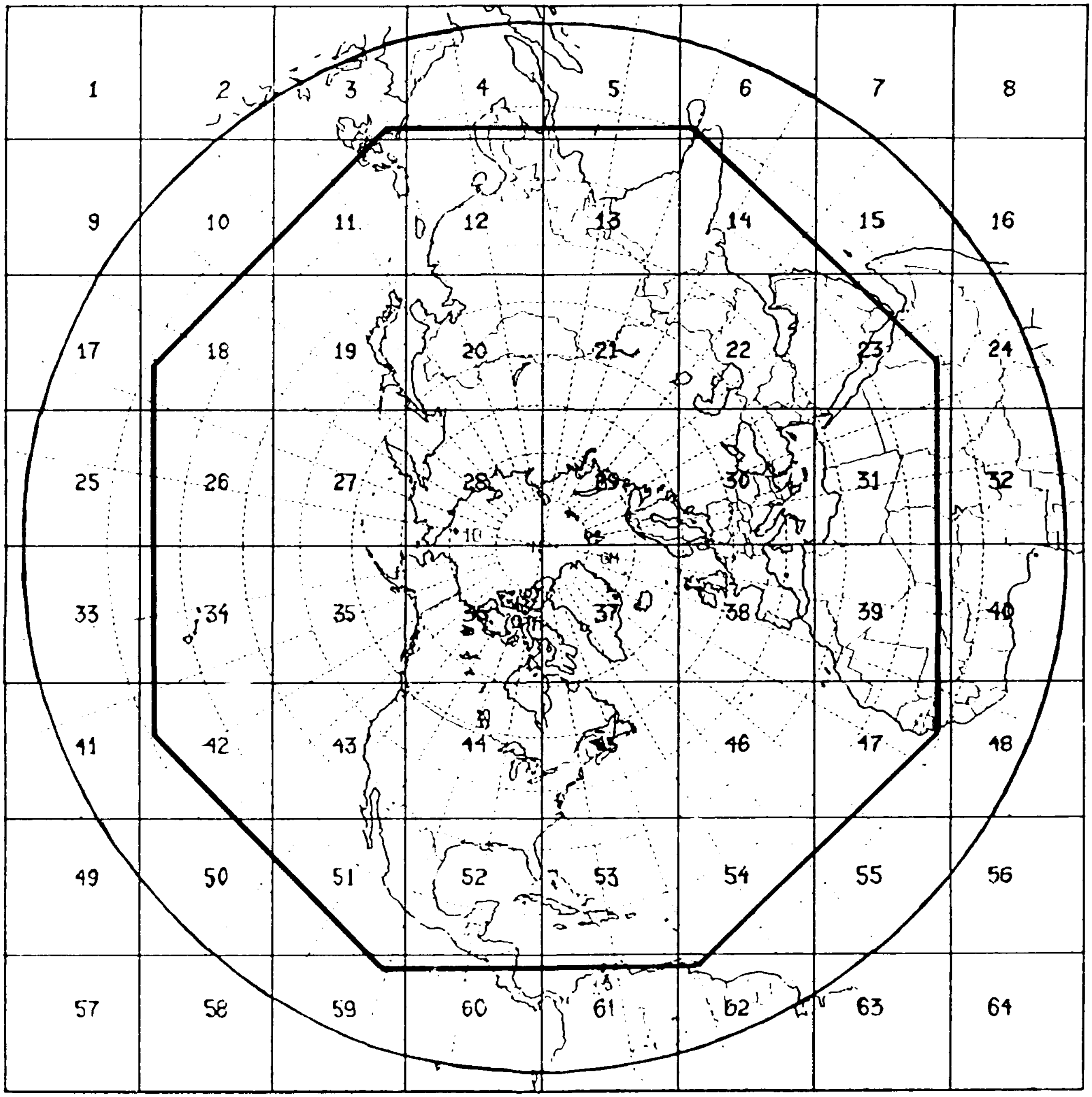


Figure 2: Illustration of the Northern Hemisphere AFGWC octagon domain.

to half-mesh, and then cloud amounts are converted to a moisture variable called condensation pressure spread, or CPS, the amount of lifting (in mb) necessary to reach saturation. Low cloud amounts are assigned high values of CPS, and high cloud amounts low values of CPS. The cloud-to-CPS curves can vary with height. Winds from a numerical forecast model are then used to define trajectories at each forecast level, and forecasts of CPS are made using a backwards trajectory technique. These CPSes are also modified through diurnal corrections and the effects of forecast vertical motion. Forecast CPS values are converted back to cloud amount.

### 2.1.2 *High-Resolution Cloud Prog*

The High-Resolution Cloud Prog, or HRCP, is the primary short-range cloud forecast model. It produces cloud forecasts at the resolution of the RTNEPH (eighth-mesh, or ~47 km at 60 degrees lat) every three hours from 0 to 9 hours. The model has a variable vertical resolution, but is generally run with four layers. The basic technique is a quasi-Langrangian trajectory scheme very similar to the 5-Layer technique. Wind trajectories are interpolated from the half-mesh winds used to drive the 5-Layer forecasts, and cloud amount is converted to CPS just as with the 5-Layer. CPS values are forecast using a backwards trajectory technique and reconverted to cloud amount. Cloud forecasts are also typically smoothed as a way of evening out the cloud frequency distribution, which tends to overforecast the occurrence of cloudy and clear conditions.

The HRCP was rewritten in 1990 to be a full trajectory model and is now substantially different from the version described in Crum (1987). The older version used a "trending technique" which was blended persistence and advection forecasts. The new version described in Kiess et al. (1993), does advective forecasts over the area covered by the 5-Layer domain and produces either a persistence or diurnal persistence forecast over the remaining tropical areas.

### 2.1.3 *TRONEW*

TRONEW is the longer-range cloud forecast model for the tropics. It is run at half-mesh, similar to 5-Layer, and is a simple diurnal persistence model; the forecast for tomorrow is today's cloud amount. Today's 18-hour forecast would thus be derived from yesterday's 6-hour old RTNEPH data. Though simple, the assumption of diurnal persistence is actually better than persistence for all but the shortest forecast lengths (approximately 6 hours or less).

#### *2.1.4 SAVDOX*

The job of the oddly named SAVDOX is to synthesize the total cloud amounts from all the forecast model data into one coherent, worldwide eighth-mesh (~48 km at 60 degrees latitude) database. All total cloud forecasts can thus be extracted from this one database. The approach to deciding which model forecast to insert is to use the most reliable forecast available; for example, a 6-hour midlatitude forecast for SAVDOX is extracted from the HRCP, although both TRONEW and 5-Layer produced less reliable forecasts for the same point. Cloud forecasts from SAVDOX are frequently examined by skilled forecasters and compared against the latest satellite data for discrepancies. If these are noted, the total cloud forecast amount can be changed in the SAVDOX database.

### *2.2 Cloud Forecast Cycle Deficiencies*

Deficiencies in the existing RTNEPH are well-documented, both in the SERCAA program plan (Snow, 1992) and the description of the existing RTNEPH (Hamill et al., 1992). These include a lack of database synopticity due to infrequent data refreshes from using DMSP data, cloud amount uncertainty, especially with low cloud and cirrus, and the tendency toward overanalysis of cloudy or clear conditions. However, the problems with the forecast models and the interaction between analysis and forecast are not well documented. Below, the problem is placed in context, and specific cloud forecast problems are explained.

#### *2.2.1 Overview*

The main current customer requirement at AFGWC is accurate cloud forecasts, with special emphasis on accurate total cloud amount. As with any forecast process, there is an inevitable decrease in the skill of forecasts with time, described pictorially in Figure 3. As shown here, skill starts off relatively high, decreases quickly at first and more slowly thereafter, eventually reaching a time where skill is nonexistent (this simple graph may not apply to TRONEW's diurnal persistence forecasts). The initial skill, as noted, is not perfect; this is due to analysis inaccuracies. An important question emerges: what is the effect of analysis accuracy on forecast skill? There are two possibilities: one is that increasing analysis skill can dramatically increase forecast skill - in effect, vertically shift the skill curve in Figure 3 upward (as illustrated in Figure 4). Another possibility of increasing forecast skill is that the positive effects will decrease quickly with time, as may happen if the forecast models are poor; this is shown in Figure 5.

Presumably the true effect of improving the analysis is somewhere between the effects shown in Figures 4 and 5, so the improvement in skill decreases with time, but not as dramatically as in Figure 5. However, the actual effect of improving the analysis skill has not been quantified. It is evident that an improved RTNEPH will improve the forecast, but if the forecast models do not have sufficient physical veracity, then analysis improvements may be wasted.

Three conclusions can be drawn from this analysis:

(1) If possible, make shortest forecast possible; if a cloud forecast for noon is needed, the skill will certainly be higher if based on a two-hour forecast using 10 AM data than a four-hour forecast using 8 AM data. AFGWC is taking steps in this direction.

(2) Obvious forecast model deficiencies should be corrected, otherwise the benefit of improving the analysis may be squandered.

(3) **The MSNEPH should be designed to translate into improved forecast skill.** If the MSNEPH is not structured to provide useful cloud information to the forecast model, then a quick dropoff in forecast skill is more likely. There may be advantages to changing the analysis variables, or conversely, wisdom hidden in the current database design. Knowing

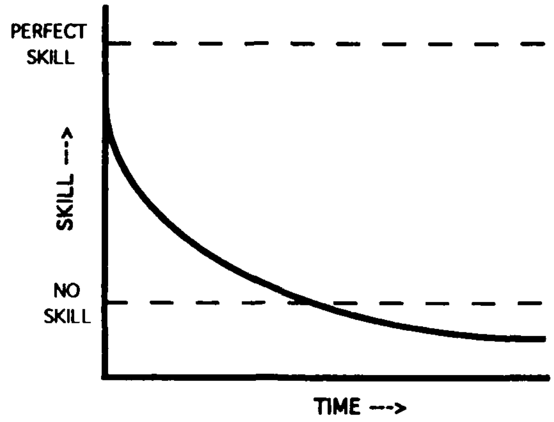


Figure 3: Plot of the typical change in forecast skill of cloud amount with time. Quality starts high, decreases rapidly at first and more slowly later on, eventually reaching a time with no demonstrable skill.

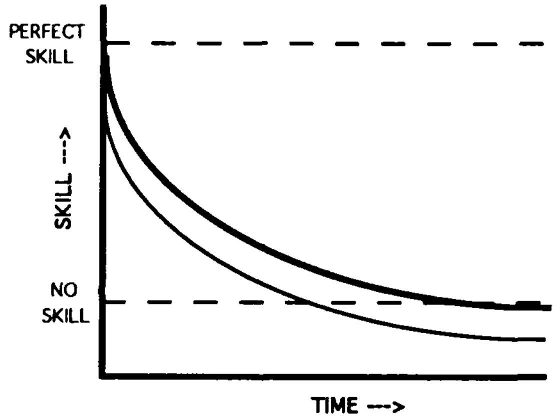


Figure 4: Hypothetical plot of the "best case" effect of an enhanced cloud analysis model on cloud forecast skill (bold line) compared with the old forecast skill curve (light line). The improvement in forecast skill is translated into improved skill at all forecast times.

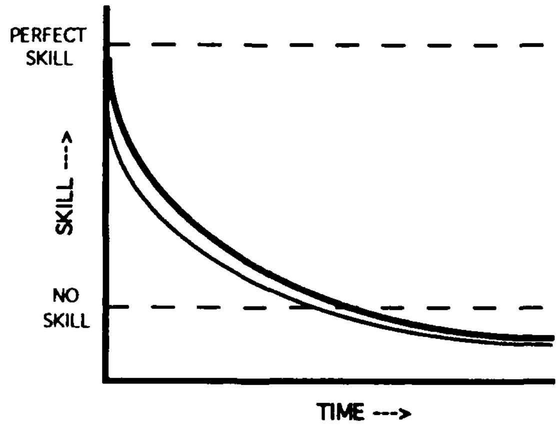


Figure 5: Hypothetical plot of the "worst case" effect of an enhanced cloud analysis model on cloud forecast skill (bold line) compared with the old forecast skill curve (light line). The improvement in forecast skill is squandered because of forecast model deficiencies.

what is right about the current procedures is as important as knowing what is wrong.

### 2.2.2 *Short-range Cloud Forecast Model Deficiencies*

It is known that there are a wide number of problems with the suite of existing cloud forecast models at AFGWC. The discussion here will focus on the deficiencies in forecasting total cloud amount, both because this parameter is the most often used, and it is the easiest to forecast and verify. Though problems are discussed here, discussion on solutions is delayed until Section 4. We start with a discussion of the short-range problems.

First let us note the strong points of the models. Experience has shown that a desirable feature of short-range cloud forecasts is that they be closely coupled to the analysis model. This is not one of the problems with the HRCP, which is closely coupled to the RTNEPH. The models run on the same grid, so there are no horizontal interpolation errors. Also, though the HRCP converts RTNEPH cloud amounts to CPS, these can be reconverted to the same cloud amount. Cloud patterns in the midlatitudes are usually coherent over many hours, so any scheme which converts the RTNEPH to information which cannot easily be converted back will usually show decreased skill. Such is the case with more conventional diagnostic cloud schemes of global numerical weather prediction models, which derive cloud forecasts from relative humidity (RH) forecasts. Since a typical RH analysis is derived from many sources (a reasonable approach for other applications), the conversion of RH to cloud amount will often yield a cloud forecast very different from the cloud analysis used as input. Thus, the basic concept of an HRCP closely tied to the cloud analysis is quite reasonable.

Despite these attractive features, many problems have been noted or implicated in working with AFGWC's short-range models. Among these are:

(1) <u>Inability to parameterize cloud development and dissipation - (HRCP)</u>. In regions where cloud changes are driven by vertical motions and not horizontal ones, the HRCP will be inaccurate. Convective clouds are especially hard to forecast. Because of this and because of software architectural considerations, AFGWC uses TRONEW's diurnal persistence forecasts

in tropical areas rather than the HRCP forecasts. However, the assumption of diurnal persistence depends on new satellite data being available frequently, for example, if there was new satellite data at 9AM yesterday but not at 12 noon, then a forecast for 12 noon today would actually be a forecast from 9AM yesterday (due to the persistence of the RTNEPH analysis in the absence of new data). Clearly, if the forecast timelines could be sufficiently shortened, then simple persistence would supply a more accurate forecast than diurnal persistence. However, results at AFGWC (not published) have shown that for forecasts beyond six hours, diurnal persistence is typically more accurate in the tropics than straight persistence.

(2) Lack of cloud advective processes - (TRONEW). Whereas the HRCP can treat the horizontal movement of clouds, the exclusive use of TRONEW in the tropics can result in poor forecasts for those cloud masses which do move horizontally, such as tropical cyclones. In a sense, the arbitrary distinction between models to treat the tropics as convective and the mid- and high-latitudes as advective may be an inappropriate simplification.

(3) Trajectory inaccuracies (HRCP).

a) Wind field errors. Horizontal advection forecasts can only be as good as the windfield used to produce the trajectories. There is potential for incorrect advection because of the many steps and resolution changes in going from the global wind forecasts to high-resolution trajectories and the numerical methods involved. In the future AFGWC is planning to remap HRCP trajectories directly from the forecast model output.

b) Vertical level assignment. If clouds are not assigned to the appropriate vertical level, even if wind forecasts are correct, the forecasts will be in error.

c) Trajectory truncation. Inaccuracies can result, even when the winds are accurate, since HRCP trajectories are arbitrarily truncated to the nearest gridpoint. Figure 6 illustrates this result with a 2-D cloud forecast. In this example, an initial gridpoint had 100 percent cloud cover but is surrounded by gridpoints with no cloud. Assuming a 20 $ms^{-1}$ wind (72 km $h^{-1}$) and a 47 km grid spacing, the cloud would move 1.53 gridpoints in the

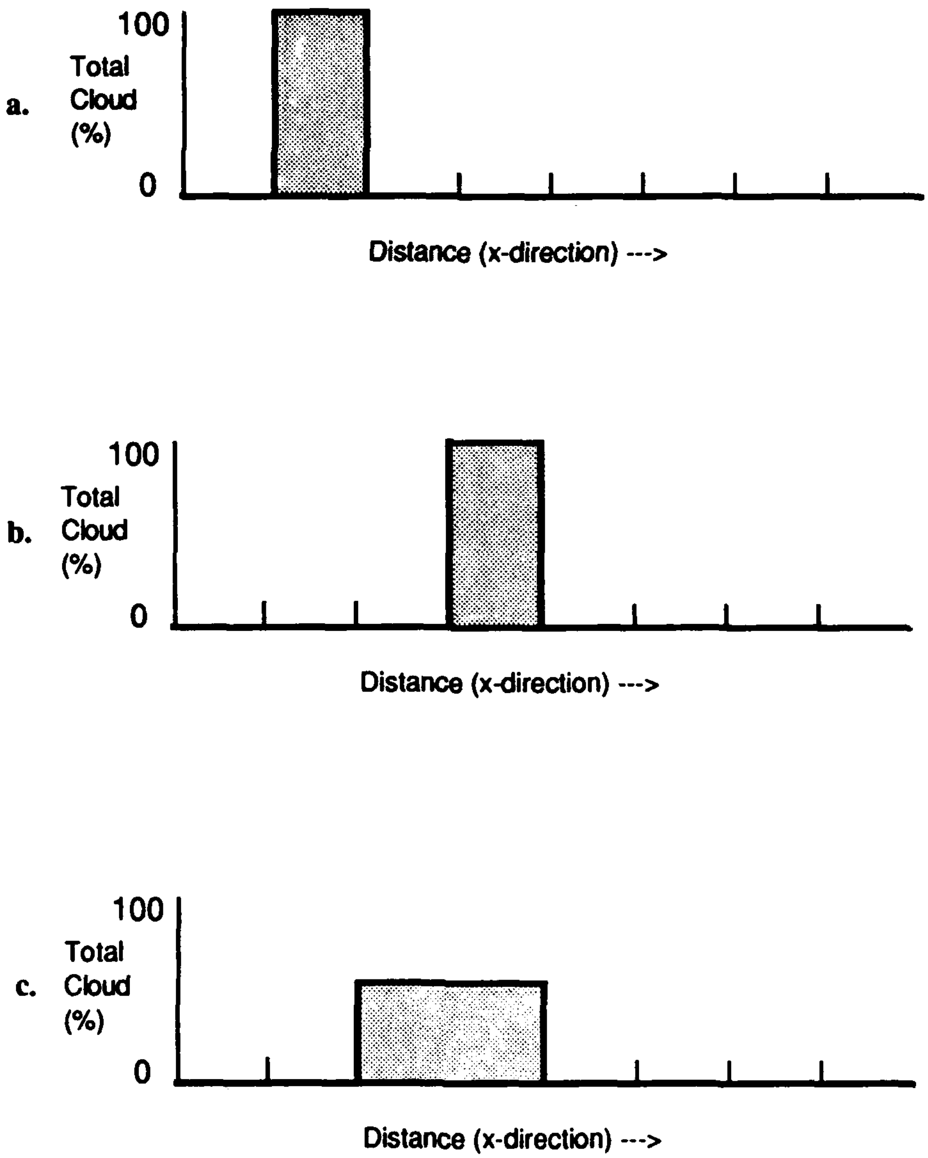


Figure 6: Illustration of how the truncation of a trajectory with a 1.53 $\Delta x$ displacement from initial condition (a) to its truncated forecast of $2\Delta x$ (b) can result in significant error. Correct forecast amounts are illustrated in (c); the initial cloud amount is distributed between two gridpoints.

HRCP's 3-hour timestep, which the model rounds off to 2.0; the forecast (neglecting later smoothing) is still a binary cloud forecast, whereas it would be appropriate to split the cloud amount almost equally between gridpoints displaced 1 and 2 points east of the original cloud location. This problem could be accentuated if the HRCP used a shorter timestep where features may be incorrectly persisted, since trajectories will have a tendency to be continuously truncated to "no displacement." For example, if the trajectories were computed every hour, and the computed displacement for a point was 0.49 gridpoints per hour, this would cause continual truncation to no displacement for every timestep, resulting in almost a 3-gridpoint error for a 6-hour forecast. However, should interpolation be used, there would be a progressive smoothing of the cloud field with time when an iterative time-marching technique is used. In this case the smoothed field from the first timestep is smoothed further at the second timestep, and so on.

d) <u>Cloud smearing</u>. In a conventional numerical prediction model, increased vertical resolution usually leads to improved forecasts. This is not necessarily true with a trajectory-based cloud forecast model. The RTNEPH often may define total cloud cover at many vertical levels due to a cloud overlap assumption built into its merge processor. For example, assume that the RTNEPH defines opaque (totally cloudy) conditions for low, mid, and high clouds. Unless the advective wind velocity is the same at all levels, the original cloud cover will be smeared out into a larger mass of opaque cloud due to the differential advective velocities. The more forecast layers in the trajectory model, the worse this problem becomes. This is a natural consequence of not having more robust physical parameterizations of cloud cover in the forecast model. Cloud cover is usually linked not only to the relative humidity but to the local forcing, which usually stays somewhat contiguous. Assuming the forcing is coincident with the cloud cover, using multiple vertical levels for advecting clouds is thus assuming that the cloud forcing is sheared as the windflow is sheared in the vertical.

### 2.2.3 *Longer-range Cloud Forecast Model Deficiencies*

AFGWC currently uses either TRONEW or 5-Layer output for forecasts over 9 hours in length. Obviously they share some of the same drawbacks discussed for the shorter-range problems, such as trajectory inaccuracies and the arbitrary separation into convective and advective forecast regimes. However, several problems with longer-range forecasts are not shared with the short-range forecasts:

(1) Outdated code. Much of the 5-Layer and TRONEW are written in outdated, poorly documented code. Though there are parameterizations for many cloud effects in 5-Layer, for example, many are not used because their precise character is unknown and unproven.

(2) Inappropriate sensitivity to initial conditions. TRONEW and 5-Layer cloud cover forecasts are closely tied to the initial cloud conditions. If the initial cloud cover changes, so will the forecasts, despite the underlying dynamics remaining the same. This strong link to the initial conditions was chosen because of the simplicity and speed of execution, but it is inappropriate except for short forecasts (say, 0 to 12 hours). Traditional numerical prediction models are less sensitive to the specified initial relative humidity or cloud cover. Forecast relative humidity and cloud cover are more closely tied to physical processes - vertical advection, condensation, etc. - simulated by the model.

### 3. Customer Requirements

This review of the requirements is not an official Air Force summary, but a synthesis of requirements as currently known.

AFGWC's cloud analysis is primarily used for the cloud forecast model initialization and verification, but there are many other users and potential users of the MSNEPH. PL/GPAS, with contractor assistance, will be surveying the meteorological climate community for input on what is desired in the cloud analysis over the long run. Some of these ideas may be implemented during the second half of the three-year SERCAA contract. Over the short run, SERCAA will be developed to meet the needs of AFGWC. However, the database design, if possible, will reflect the broader needs of other Air Force and civilian users. Below, we first set up a matrix of the expected user priorities, and then discuss in depth the AFGWC forecast requirements and the unresolved issues.

## *3.1 Matrix of User Requirements*

### *3.1.1 Customers*

For purpose of discussion, let us arbitrarily separate the user community into three blocks:

(1) 1-1 Customer. AFGWC's existing customer is their Air Force Precedence 1-1 strategic customer. In-depth requirements of this customer will be described in Section 3.2.

(2) Tactical User. This user does not extensively use AFGWC products now, but in the future, with operational usage of the Battlefield Weather Observing and Forecast System (BWOFS) and the Combat Weather System (CWS), AFGWC cloud products may be more in demand by in-the-field users. These products would be used on-site to improve aviation cloud forecasts and for the calculation of tactical decision aids.

(3) Civilian/Climate User. Into this catch-all category are lumped all other users; these may be university researchers verifying model forecasts, weapons designers seeking cloud climatology information, or many other possible users. We will assume the primary users are climate modelers.

### *3.1.2 Matrix of Users and Requirements*

Without discussing specifics, we will assume a cloud analysis and forecast is available to each of the users. Further, we will assume that layered and total cloud amount, layer type, cloud base and top elevation, and emissivity are both analyzed and forecasted (it is understood that not all parameters will be analyzed or forecasted with the same reliability). Table 1 gives a simple matrix of the relative priority each customer is expected to assign to each parameter:

| Analysis / Fcst Parameter | Military 1-1 Customer | Military tactical user | Civilian / climate user |
|---|---|---|---|
| **Analysis:** | | | |
| Total cloud | HI | MOD | HI |
| Layer amount | MOD | HI | MOD |
| Layer type | MOD | MOD | MOD |
| Emissivity | MOD | MOD | HI |
| Cloud top height | MOD | MOD | LOW |
| Cloud base height | MOD | HI | LOW |
| **Forecast:** | | | |
| Total cloud | HI | MOD | NONE |
| Layer amount | LOW | HI | NONE |
| Layer type | LOW | MOD | NONE |
| Emissivity | LOW | MOD | NONE |
| Cloud top height | MOD | MOD | NONE |
| Cloud base height | LOW | HI | NONE |

Table 1: Matrix of the priority each of the three customers of AFGWC cloud products will assign to each of the available forecast products.

The 1-1 customer is primarily interested in total cloud cover forecasts. The analysis parameters are necessary for model verification and initialization, however, if it were possible to achieve an accurate forecast without specific analysis parameters, such as cloud-top height, this customer would be unaffected. Cloud-top height in the analysis was rated as a moderate, not because of direct demands by the customer but because of the necessity of this information for initializing the cloud forecast model accurately. A point analysis model for calculating atmospheric effects is also run which is highly sensitive to the full suite of RTNEPH layered cloud parameters, thus, these parameters were listed as moderate in importance.

Drawing on the lessons from Desert Storm, the tactical user is most concerned about having accurate forecasts of obscurations to vision for himself and his weapons. Cloud-free line-of-sight calculations will be derived from the forecast parameters, and these calculations are very sensitive to the layered cloud structure. A particularly crucial parameter is the cloud base height or ceiling height, since this may determine whether a pilot can fly high enough to lock on to ground targets without being exposed to ground-based return fire. Corresponding analysis parameters were assigned a similar priority presuming that the forecast of each element could be no more accurate than the analysis.

For the civilian, off-line user, cloud forecasts are unimportant, but analysis parameters are important. The analysis parameter's importance may vary from user to user, but most are expected to desire an accurate total cloud cover forecast and accurate layer information which would improve radiative transfer calculations.

Meshing these requirements together, we assume the most important element is analysis and forecast total cloud. The next most important element(s) are the forecast layer structure, with the 1-1 customer requiring cloud-top height (indirectly) for the model initialization and the tactical user focusing on cloud amount and base height. Since the RTNEPH is primarily a satellite-derived nephanalysis, obscuring upper decks of cloud will frequently prevent accurate determination of lower cloud layer characteristics. These properties can only be inferred at best, and the use of inferences rather than actual data in creating the database may preclude its usefulness for other users such as climate researchers. SERCAA designers will need Air Force guidance on such issues before algorithm development starts.

### 3.2 *Requirements for the MSNEPH*

At the first SERCAA technical exchange meeting, the gross database structure for the MSNEPH was outlined by AFGWC. This database structure is presumably the structure of choice for cloud forecast model initialization. The requirements are as follows:

a) Coverage: Global. Note however that MSNEPH testing under SERCAA will initially be limited to a smaller area, primarily over North and Central America, where data is readily available.

b) Resolution:

(1) Horizontal: approximately 1/16th mesh (25 km). The map projection has not been decided yet. This requirement is driven by the

customer requirement for cloud forecasts at this resolution. This requirement does not specify, however, that input satellite imagery must be mapped to this projection before use in the cloud analysis.

(2) Vertical: includes at least three cloud levels, low, mid, and high, defined as "floating layers" as is done with the current RTNEPH. It is assumed that there is inadequate information from satellites to define more than three distinct levels.

c) Input data: Raw satellite imagery from DMSP, GOES (Geostationary Operational Environmental Satellite), NOAA/AVHRR (National Oceanic and Atmospheric Administration - Advanced Very High Resolution Radiometer), and foreign geostationary satellites are to be processed. To support this, it is assumed that all existing and planned AFGWC databases will still be available, including:

(1) "Best reports" database of conventional cloud data.

(2) Surface/skin temperature analysis and forecast (including sea surface temperatures).

(3) Snow and ice analysis.

(4) Analysis and forecast fields of temperature, geopotential height, and relative or specific humidity.

(5) Gridded and raw SSM/I (DMSP's Microwave Imager) data.

(6) Gridded and raw SSM/T-2 (DMSP's Microwave Moisture Sounder) data.

d) Output database Parameters. The MSNEPH will output at least the following:

(1) Total cloud amount.

(2) Layered cloud amount.

(3) Layer cloud top altitude (MSL).

(4) Layer cloud base altitude (MSL).

(5) Cloud layer type. Permissible cloud types will include but not be limited to: cirrus, cirrostratus, cirrocumulus, altostratus, altocumulus, cumulus, stratocumulus, stratus, fog, cumulonimbus, and clear.

(6) Diagnostic information such as:

- Satellite data used (e.g. DMSP satellite F11, GOES-Next).

- Time of most recent satellite data.
- Analysis quality. This is yet to be precisely defined, but will quantify the presumed accuracy of the analysis at that point.
- Thin cloud (for cirrus).

e) Timeliness: Not presently defined for the MSNEPH, but when transferred to AFGWC's global operations, the software will need to be able to produce a worldwide cloud depiction every hour, analyzing and synthesizing all the latest satellite data. We assume that the MSNEPH will help define the sizing of the host computer system, and not vice versa.

### *3.3 Unresolved Issues*

Though the overall structure of the database is clearly specified, some of the important internal algorithmic decisions have not yet been made. This includes:

a) Bogusing. This refers to the modification of the RTNEPH cloud analysis by trained forecasters. Currently the RTNEPH boguses override the existing cloud analysis. The reason for bogusing the RTNEPH is required since the analysis is often inaccurate. For the MSNEPH, extensive bogusing may no longer be desirable, for the following reasons:

(1) The MSNEPH should be more accurate in general.

(2) The short amount of time AFGWC has to produce and disseminate cloud forecasts may not allow extensive bogusing, yet the seeming need for bogusing will increase with increased processing of satellite data.

(3) Bogusing the MSNEPH includes subjective judgment of cloud amount which may degrade the usefulness of the MSNEPH to climate users and weapon developers requiring database consistency.

Currently the thinking at AFGWC is to allow bogusing of the MSNEPH with AFGWC forecasters being alerted to problem areas in the MSNEPH through its internal quality flag. It is possible that the MSNEPH database will archive the pre-bogused cloud statistics, or that the boguses can be made to affect the forecast model initialization without affecting the cloud analysis database. With the cloud forecast model output being bogused, either the analysis or forecast bogus may be an extraneous step. Thus, the following questions emerge:

(1) Will the RTNEPH continue to be bogused?

(2) If so, must the MSNEPH design for cloud analysis integration account for previously bogused areas, as does the current RTNEPH? The current RTNEPH will not allow a new satellite analysis to overwrite a bogused area for a few hours.

b) Database stability. Should the MSNEPH be designed to allow interactive tuning of the cloud analysis or as a long-term, stable system? By allowing quick tuning of the MSNEPH, the analysis may be adjusted to locally or globally increase or decrease cloud cover, depending on the perceived bias of the model. Although the ability to tune the models will continue at AFGWC, we believe it is essential that the MSNEPH be designed to eliminate or at least reduce human tuning, for the following reasons:

(1) Tuning is an inherently subjective process, often done looking at a limited amount of data rather than a time series of many weeks or months. Thus, tuning the analysis may fix today's problems yet detract from the quality of the analysis days later when weather regimes change.

(2) Planned design for the satellite data processing algorithms will obviate the need for extensive human intervention, and design for the integration algorithms will require a database with stable error characteristics. Tuning of the model without recalculation of the associated error characteristics can degrade the end product rather than improve it.

(3) If designed for stability, the MSNEPH would be more useful to a wide variety of users in the weapons planning and climate community. These users, often seeking month-to-month or year-to-year comparisons of

cloud cover, would use the database only if it used a consistent algorithm throughout.

(4) Any future statistically based cloud forecast models would be inherently less accurate if the quality of the MSNEPH is subject to the vagaries of a frequently tuned nephanalysis.

c) Conventional data. The RTNEPH database currently uses and will continue to use (refer p. 41) conventional cloud observations to augment the satellite-derived nephanalysis. The presumed benefit of this data is the determination of cloud cover in the absence of new satellite data or the detection of low cloud when the satellite data indicates an obscuring upper cloud deck. Including this data may be particularly useful to the tactical user of the MSNEPH, with important requirements for accurate low cloud amount and height. Conversely, many conventional cloud observations are unreliable, especially at night and in the Eastern European group of nations. Furthermore, conventional data cloud amounts are often coded as only clear, scattered, broken, or overcast, which, if coded as 0, 25, 75, and 100 percent cloud will guarantee significant error in the estimate of total cloud. Lastly, the climate community users may prefer a pure satellite database rather than a mixed-source database. Thus, there is an outstanding decision on whether the MSNEPH should continue to use conventional cloud observations, and if so, when and how. Perhaps conventional observations can be more appropriately used during the bogusing process as a visual aid for identifying areas with incorrect satellite-derived analyses.

d) Cloud overlap assumptions. As described in 2.2.2 (4), the HRCP cloud forecast model often smears out the analyzed cloud stacked in layers at one gridpoint into a wider area of multi-deck clouds. This effect can be traced back to the random cloud overlap assumption built into the RTNEPH merge processor. In the case of a satellite analysis determining two layers of cloud (with the sum of the layer amounts less than 100%), the lower layer amount is increased to account for viewing obscuration by the upper layer. The amount (I) to increase the lower layer amount is determined by:

$$I = (CL*CA) / (100 - CA)$$

where CL is the current layer amount and CA the upper layer amount. CL is readjusted then to be CL+I.

This inference of lower level cloud amount no doubt affects the quality of the HRCP. The model has never been tested with the RTNEPH overlap assumption turned off. It is possible that a more accurate forecast would result. Conversely, if these sort of assumptions are not built into the MSNEPH, the database will reflect only the satellite-sensed cloud cover and will drastically underestimate mid- and low-cloud amounts. Again, the climate community may prefer a database where only observed cloud is stored, not inferred cloud, but the tactical user of the MSNEPH may be badly affected by the underestimate of low cloud. The design of a new Merge processor (i.e., cloud analysis integration algorithm) may not use exactly the same cloud overlap assumptions, but the question is whether any inference of lower cloud amount should be made, or whether only observed cloud amounts should be used. If HRCP accuracy is the crucial, driving requirement, testing of the relative accuracy with and without the overlap assumption may be necessary before making this decision. Perhaps the wise choice is to archive only what is sensed and have each database user enhance the baseline database with whatever overlap assumptions are valid for their particular application.

e) Augmented parameters. Several additional database parameters such as cloud emissivity, particle size distribution, and water content have been suggested for inclusion in the MSNEPH database. Such parameters can be derived from some satellites and not from others. These parameters could potentially be part of the final MSNEPH database or could be stored into a separate database unique to the originating satellite. If part of the MSNEPH database, the parameters would also have to be filtered through the new analysis integration processor and checked for consistency against the other analysis parameters. For example, if the merge processor indicates no cloud, nonzero cloud liquid water should not be allowed. If such augmented parameters are likely to be a part of the MSNEPH at some future point, even if not included now, future inclusion would be easier if some provisions were made now for such parameters.

f) Parameters not included in the MSNEPH. Design requirements did not specify that the MSNEPH archive weather and visibility from observations, as it does currently. This information is readily available from other databases.

## 4. Discussion of Cloud Forecast Algorithms

The deficiencies of the existing AFGWC cloud forecast schemes were documented in Section 2. AFGWC, or possibly a contractor selected by AFGWC, will be upgrading the cloud forecast models to hand off to the computer system contractor for coding (presumably in ADA) and implementation. There are a myriad of potential improvements which can be made to the cloud forecast models for both short- and longer-range forecasts. Below, we briefly review the possible algorithms and the important analysis parameters necessary to initialize these forecasts. Potentially the initialization requirements of these methods may impose additional requirements on the MSNEPH database.

### *4.1 Short-range Forecast Algorithm Improvements*

AFGWC closely ties its short-range cloud forecasts to the RTNEPH. This assumption will, if anything, become more appropriate, not less, as the MSNEPH is coded to produce a truly synoptic database and the cloud forecasts needed are shorter in length. However, improvements to the forecast models are likely to be made in the future. Such improvements might include:

a) Synthesis of advective and convective algorithms. Rather than arbitrarily distinguishing between using an advective forecast algorithm at mid- and high-latitudes and a diurnal persistence model at low latitudes, it may be profitable to combine elements of both into one short-range forecast model. In this way, whatever convective cloud parameterization is used may be applied more broadly, such as to mid-latitude summers. Similarly, advective processes would not be neglected in the tropics.

b) Parameterization of diurnal effects. A diurnal persistence forecast model may indeed prove nearly unbeatable as a forecast algorithm for tropical forecasts in the six- to nine-hour range; this may be particularly true with the MSNEPH and its more rapid refresh. However, AFGWC may shorten its cloud forecast cycle, thereby allowing shorter-length forecasts to be used (e.g., a 3-hour forecast instead of a 5-hour forecast). In this case, diurnal persistence may show less skill than straight persistence. However, it is reasonable in this case to assume that additional forecast skill might be gained by including a forecast of the short-term development and dissipation of cloud. Some of the possibilities here include:

(1) Diurnal trending: Assume a short-term cloud forecast based on persistence or advection is available but does not include any diurnal effects ($F_{no\_d1}$). A simple forecast with diurnal parameterization ($F_{d1}$) would be:

$$F_{d1} = F_{no\_d1} + D_y$$

where $D_y$ refers to the change in cloud amount yesterday at the forecast time and analysis time, respectively, or could represent the error in the advective forecast solution the previous day. This method was tried briefly at AFGWC with little success in 1988, but the lack of success may have been due to the poor analysis quality rather than deficiencies in the approach.

(2) Statistical approach. A MOS, Perfect Prog, or neural network approach using the MSNEPH data may show potential for improving forecasts of diurnal effects. A multivariate approach could include not only cloud analysis information as predictors, but also other factors such as moisture convergence and stability as supplied from a numerical forecast model. One disadvantage to this approach is the extensive training data sets necessary (perfect prog excepted). These datasets presumably will not be available until after the prototype MSNEPH is available, which may be too late to use in time for the computer upgrade. Another disadvantage is the tendency for regression-based techniques to drive forecast amounts toward partly cloudy to a greater extent than is observed. Ideally, a chosen statistical approach would be

constrained to produce a cloud frequency distribution spanning the range of cloud amounts rather than being clustered in the middle amounts.

(3) Dynamic approach. A boundary layer forecast model such as the OSU model (Ek and Mahrt, 1989) may be useful for predicting the short-term evolution of low-level cloud cover due to the changes in stability. There is already a version of the OSU boundary layer model at AFGWC which could be enhanced with minimal cost. A potential downside of this technique is the known sensitivity of such models to initial conditions such as surface moisture, which, if not correctly analyzed, can dramatically affect the accuracy of the forecast fields.

c) Correlations-based cloud forecasts. This is a rather simple cloud forecast technique tailored to geostationary satellite data. It is a trajectory-based technique, but trajectories are derived through a cross-correlations analysis as described in Hamill and Nehrkorn (1993). The correlations forecasts work with individual satellite pixels, not compacted cloud analyses, so a coarser-mesh cloud forecast would have to be generated from the pixel-by-pixel cloud forecast generated from the correlations scheme. The scheme has the advantage of demonstrated accuracy (10-30 percent improvement over persistence for 1/2 to 2 1/2-hour forecasts). Also, by virtue of using a one-level windfield, it does not "smear" the RTNEPH layered cloud amounts as does the existing HRCP. Its primary disadvantage is its considerable algorithmic difference from the existing cloud forecast structure at AFGWC. This correlations approach requires the nephanalysis function and layer structure be determined after the forecast, not before.

d) One-level HRCP. Work with a cross-correlations forecast scheme showed the potential benefit of using a single-level cloud forecast scheme. Such a forecast scheme does not have a positive cloud amount bias in the forecasts caused by the "smearing" effect previously discussed. It may be possible to apply cross-correlations technology to the HRCP. The algorithm would advect clouds with a one-level windfield derived from multi-level GSM data and locally warped to the predominant cloud elevation. Thus, areas with mostly low cloud might use the 850 mb or gradient winds, and a nearby jet cirrus would use the 300 mb winds, with intermediate winds in between.

AFGWC has recently tried some simple one-level HRCP experiments, advecting all clouds with one-level winds (500 or 700 mb). This scheme has showed less skill than a multi-level model. However, the wind field for this test was a one-level windfield, not locally warped to the appropriate elevation, as is done with the cross-correlations. Thus, there is still a possibility that a well-engineered one-level model may perform better than the existing HRCP.

### *4.2 Long-range Forecast Algorithm Improvements.*

AFGWC continues to rely on TRONEW and 5-Layer for its longer-range cloud forecasts. Whereas for short forecasts a sensitivity to initial cloud conditions is desirable, for longer-range forecasts it is less appropriate, as dynamic effects increase in importance. Other possible replacement schemes exist:

a) Improved trajectory methods. A larger, faster host computer would allow the resolution of the forecast model to be improved so that potentially a high-resolution, HRCP-type forecast could be run for as long as the 5-Layer is currently run. Whether the improved resolution would lead to an improved forecast is not yet known. Presumably the model would be more accurate due to smaller trajectory errors with the finer grid.

b) Diagnostic schemes. These schemes are so named because cloud cover is diagnosed from forecasts of the other state variables rather than being treated as a forecast variable itself. Examples of this scheme include the Slingo scheme (Slingo, 1987), used operationally at ECMWF, NMC, and in the Community Climate Model at NCAR, and the cloud-curve algorithm, or CCA (Mitchell and Hahn, 1989). Both have the advantage of being computationally efficient, and because the diagnostic scheme does not feed back to the rest of the forecast variables, this scheme can be run off-line, or even at a different location. This would be a tremendous advantage for AFGWC if plans require the relocation of the global forecast capability to another center, such as the Navy's. These schemes have some notable disadvantages, however. One is that they are limited by the forecast accuracy of the driving NWP model; if this model does not forecast relative humidity and vertical motion accurately, then the cloud forecasts will certainly be poor. Presumably, this is a primary reason for the lackluster forecast skill exhibited by the CCA scheme when tested with AFGWC's forecast model (Trapnell, 1992). Another disadvantage is this model will not produce forecasts competitive with the HRCP for short timescales, hence requiring at least two forecast models, one for the short, and one for the long-range.

There are a number of possible variants AFGWC may wish to try with the diagnostic cloud forecast schemes. Among these are:

(1) CCA. The cloud-curve algorithm is a simple cloud forecast scheme which can correct for relative humidity biases in the forecast. Forecast relative humidities are converted to cloud amount based upon a climatological relationship between frequency distributions of RTNEPH cloud amount and RH. Forecast skill may be improved by separating the cloud forecast generation into climatological regions so that different RH-to-cloud amount curves are used for different regimes. The downside of this separation, however, are possible odd discontinuities in the analysis at regime boundaries. The CCA scheme or its successor could be applied to work with AFGWC, Navy, or other forecast model output. However, when coupled with AFGWC's forecast model, this scheme was less skillful than the existing 5-Layer. Improvements to the scheme or to the driving NWP model are clearly necessary before AFGWC should consider replacing the 5-Layer.

(2) Slingo scheme. This technique relates cloud amount not only to relative humidity but also to forecast parameters such as convective activity and vertical stability. This scheme cannot be applied to AFGWC's forecast model with its limited convective and boundary layer parameterization, but if Navy or National Weather Service forecast data is available, AFGWC could postprocess the forecast fields using this scheme. A disadvantage of this scheme is its use of a fixed relative humidity-to-cloud relationship which cannot account for model biases dependent on time or location.

(3) Hybrid schemes. A blend of the strengths of the CCA and Slingo schemes may provide better forecast skill than either individually. NMC has done some preliminary exploration on how to improve their cloud forecast schemes through a hybridization (personal communication, Ken Mitchell).

(4) Synoptic climatology. Zivkovic and Louis (1992) outline a procedure for determining meteorologically similar regions. This scheme filters analysis or forecast data into a principal component analysis resulting in a definition of climatic regime based on temperature, stability, humidity, etc. A unique cloud cover representative of the mean may be assigned to each regime, or a unique cloud forecast scheme developed within the regime. Whereas the CCA regimes are strictly a function of position on the globe, use of a synoptic climatology approach would allow a given location's cloud amount relationship to be determined differently depending on the forecast air mass. This scheme was more accurate than the Geleyn cloud scheme (the predecessor of the Slingo scheme) in a simple test with analysis data. This relatively new scheme has not been fully tested, however. It could suffer some of the same discontinuities as the CCA algorithm at the boundaries of regimes, though presumably these regime boundaries would more closely reflect air mass boundaries.

c) Prognostic schemes. Several meso-alpha scale numerical forecast models now treat cloud liquid water (and ice) explicitly as a forecast variable (e.g., Sundqvist et al., 1989). These schemes are referred to as prognostic schemes. The potential accuracy of these schemes makes them quite attractive, but they can be computationally very expensive, and are certainly unproven and risky right now.

### *4.3 Analysis Parameters Required for Forecast Model Initialization*

With many different cloud forecast schemes, we can expect that many different analysis parameters might be necessary for initialization or verification. The MSNEPH database must certainly include all the parameters that the planned suite of AFGWC forecast models need, supplied at the resolution of the forecast model. Here is how we perceive the importance of each database parameter in relation to the interaction with the cloud forecast models.

a) Total cloud amount: This is the most important parameter of all. Since total cloud amount forecasts are the most critical parameter, the analysis of total cloud is necessary for both forecast initialization and verification. Lacking any cloud forecast scheme, a persisted total cloud forecast may be used as the cloud amount forecast, and for very short-range forecasts (say, 0-2 hour), most forecast schemes are unlikely to perform more skillfully than persistence anyway.

b) Layer cloud-top height: This is judged to be the next-most important parameter because it will be used for determining the level of winds used in the advective forecasts. Specifying the top level cloud, which is most readily visible from satellite, is presumed to be more important than specifying the height of the obscured layers. Lacking other parameters such as thickness, base, and layer structure of the lower-layer clouds, a simple forecast model could be run using just the total cloud amount and layer height - a simple, one-level cloud forecast model similar to the cross-correlations approach (Hamill and Nehrkorn, 1993) with the advective wind velocity field locally warped to the top layer height. Thus, next to determining the total cloud accurately, the accuracy of this parameter must be emphasized in the integration algorithm.

c) Layer cloud-base height (and thickness): These parameters are more important for the tactical user than the strategic user concerned with accurate short-range total cloud forecasts. Because it is difficult to estimate cloud base accurately from satellite and because it is possible to design a robust cloud forecast scheme of total cloud without a precise definition of cloud base, accurate specification of this parameter will not be a primary focus during the initial stages of the integration algorithm design. Techniques for using conventional data for cloud base height, however, are readily available, and will be discussed in the next chapter.

d) Layer emissivity and other augmented parameters: There is currently no listed requirement for augmented cloud parameters such as emissivity, but such parameters are readily derived from AVHRR and the present GOES data and thus could be produced and stored into the final database. Because of the short timelines to producing a prototype cloud analysis integration algorithm and the lack of a clear requirement for this parameter, this will be deemphasized during the initial phase of SERCAA.

## 5. MSNEPH Cloud Analysis Integration and Forecast Model Interaction

Many of the MSNEPH database requirements have already been defined and the algorithms to be used for the raw satellite data analysis are well-explored. However, an integral part of the MSNEPH will be the integration of the varied satellite analyses into the final analysis database. This technology has not yet been explored or defined, so we first will step back here for a look at the big picture of the interaction of analysis and forecast models for the purpose of exploring a possible optimum design. We then outline in more detail where we expect the cloud analysis integration research may be headed under SERCAA.

### *5.1 System Design Options*

We now outline a variety of design options for the MSNEPH and forecast models, and discuss the benefits and drawbacks of each approach.

#### *5.1.1 Algorithm 1: Full MSNEPH Cloud Integration*

(1) Description: Let us assume that many different algorithms are available for processing the same satellite data. For example, DMSP data could be examined in a spatial coherence technique (Coakley and Bretherton, 1982), through a single-channel IR method, etc. Each of these would write out the pertinent output parameters to their own separate database. The data integration algorithm would then synthesize these independent cloud analyses into one coherent database, which would feed the cloud forecast models. This process is illustrated pictorially in Figure 7.

(2) Advantages: Such a scheme could potentially perform cloud analysis integration in its most robust form; all the various benefits and drawbacks of each algorithm could be weighed against all the others, arriving at a truly optimum, lowest-error cloud analysis.

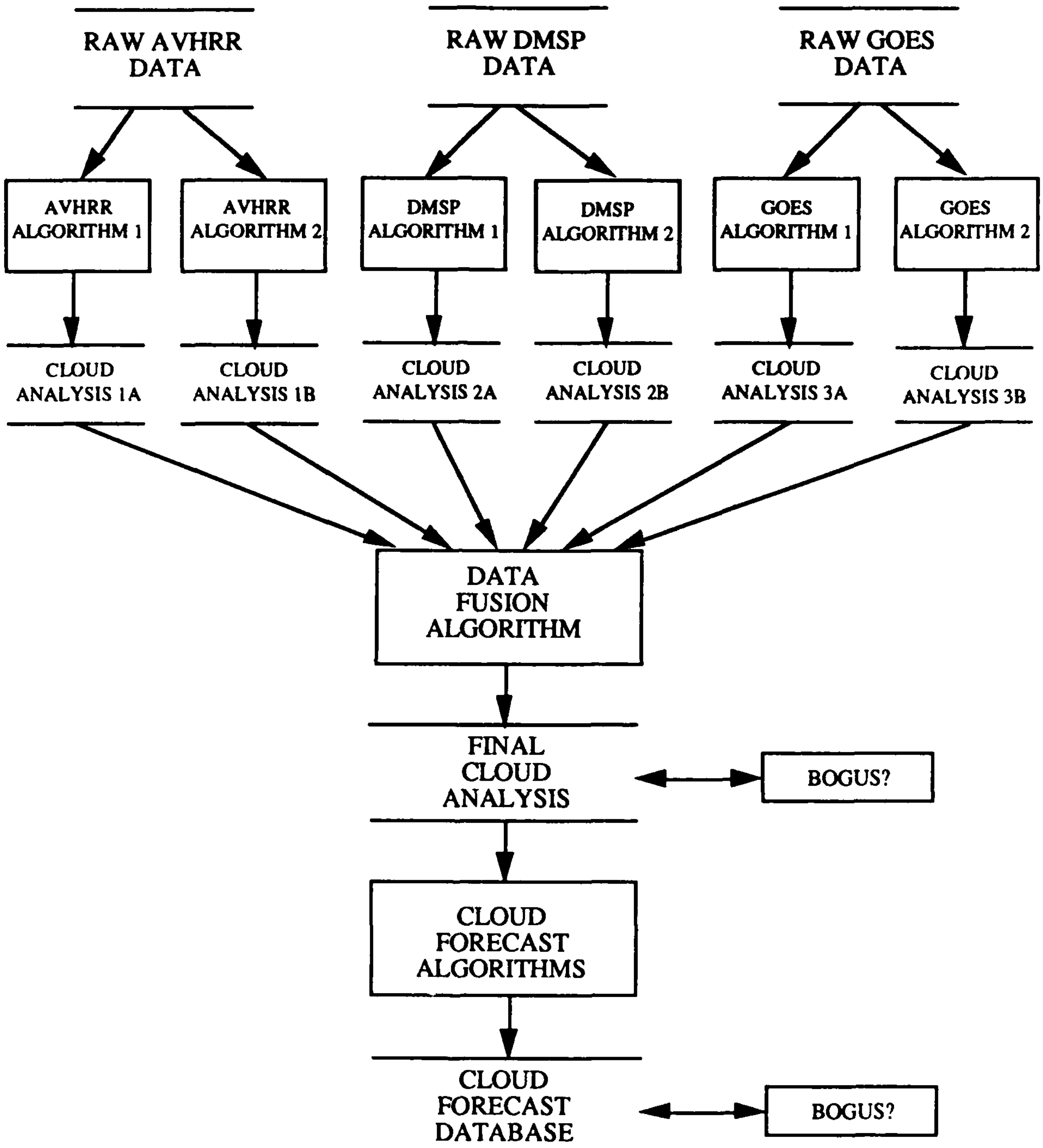


Figure 7: Illustration of the data and processing flow for the "full cloud analysis integration" algorithm.

(3) Disadvantages: The data integration algorithm may be too complex to design in a short amount of time. The more pieces of information feeding into the algorithm, the greater the number of possible combinations that will need to be consistency-checked against each other. The design could potentially become unwieldy and future code modifications very difficult.

### *5.1.2 Algorithm 2: Stepwise MSNEPH Cloud Integration*

(1) Description: We again assume many possible data analysis algorithms are available for each type of imagery. In stepwise cloud integration, however, for each satellite a preliminary fusion algorithm is run to internally resolve analysis inconsistencies. One final satellite analysis is produced for each sensor. Thus, if both a hybrid bispectral threshold method and spatial coherence technique were run for the DMSP, the two analyses would be combined into one before integration with AVHRR and GOES. The job of the integration algorithm is thus simplified (compared with algorithm 1); the module again produces a final cloud analysis used to initialize the forecast models. This process is illustrated in Figure 8.

(2) Advantages: Like the first algorithm, the cloud analysis integration would be able to weigh all the various benefits and drawbacks of each algorithm against all the others. The complexity of the integration algorithm would be reduced, with only three analyses to intercompare. This algorithm also follows the general plan of the TACNEPH program, which produces consistent cloud analyses for each sensor. Thus, the transition of TACNEPH satellite technology to MSNEPH technology would be easier.

(3) Disadvantages: The accuracy may potentially be reduced from algorithm 1, and though the complexity of the integration algorithm would be reduced for each individual part, there would in a sense now be four smaller integration algorithms rather than one big one; thus, the complexity does not totally go away, and thus may still be difficult to fully design and implement in a short amount of time.

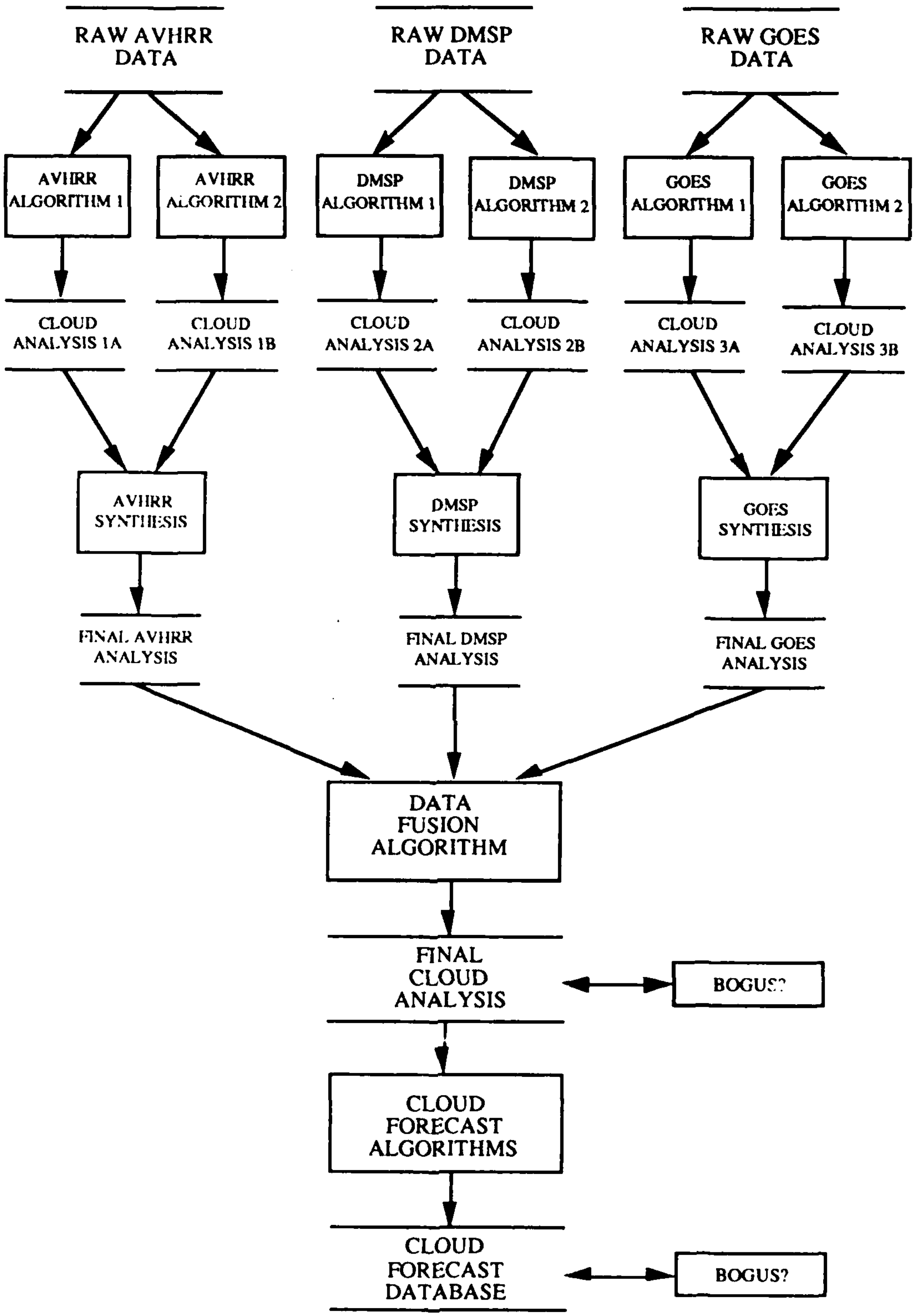


Figure 8: Illustration of the data and processing flow for the "stepwise cloud analysis integration" algorithm.

### *5.1.3 Algorithm 3: Extension of Current RTNEPH Design*

(1) Description: Currently the RTNEPH has one satellite cloud analysis, with the analysis corresponding to the most recent satellite data available for a given area. The merge processor of the current RTNEPH could be used almost intact, writing out a final cloud analysis which would feed the forecast algorithms. The process is illustrated pictorially in Figure 9.

(2) Advantages: Presumably the MSNEPH would be more accurate than the current RTNEPH despite a robust integration algorithm, since the individual satellite analyses would be improved with new algorithms and able to process new data sources. The risk level would be low, with limited new technology, and thus greater assurance of having a workable cloud system available for delivery to the contractor at the specified time.

(3) Disadvantages: The cloud analysis is likely to be less accurate than with either of the previous algorithms since strengths and weaknesses of the various satellite algorithms would not be intercompared and resolved in a cloud analysis integration step.

### *5.1.4 Algorithm 4: Forecast Integration*

(1) Description: Rather than having the MSNEPH perform the cloud analysis integration, it is also possible to run separate cloud forecasts unique to each satellite and combine the forecast output rather than the analysis output. This is illustrated in Figure 10.

(2) Advantages: The climate community users may actually prefer separate analysis databases for each satellite rather than one communal satellite database; in this way they know exactly what they are getting. Also, the current HRCP design, which produces cloud forecasts only a quarter-orbit at a time, could be preserved, and the correlations methodology could be used for the geostationary cloud forecasts. Potentially this could result in fewer software modifications necessary before the computer upgrade, making the transition less risky.

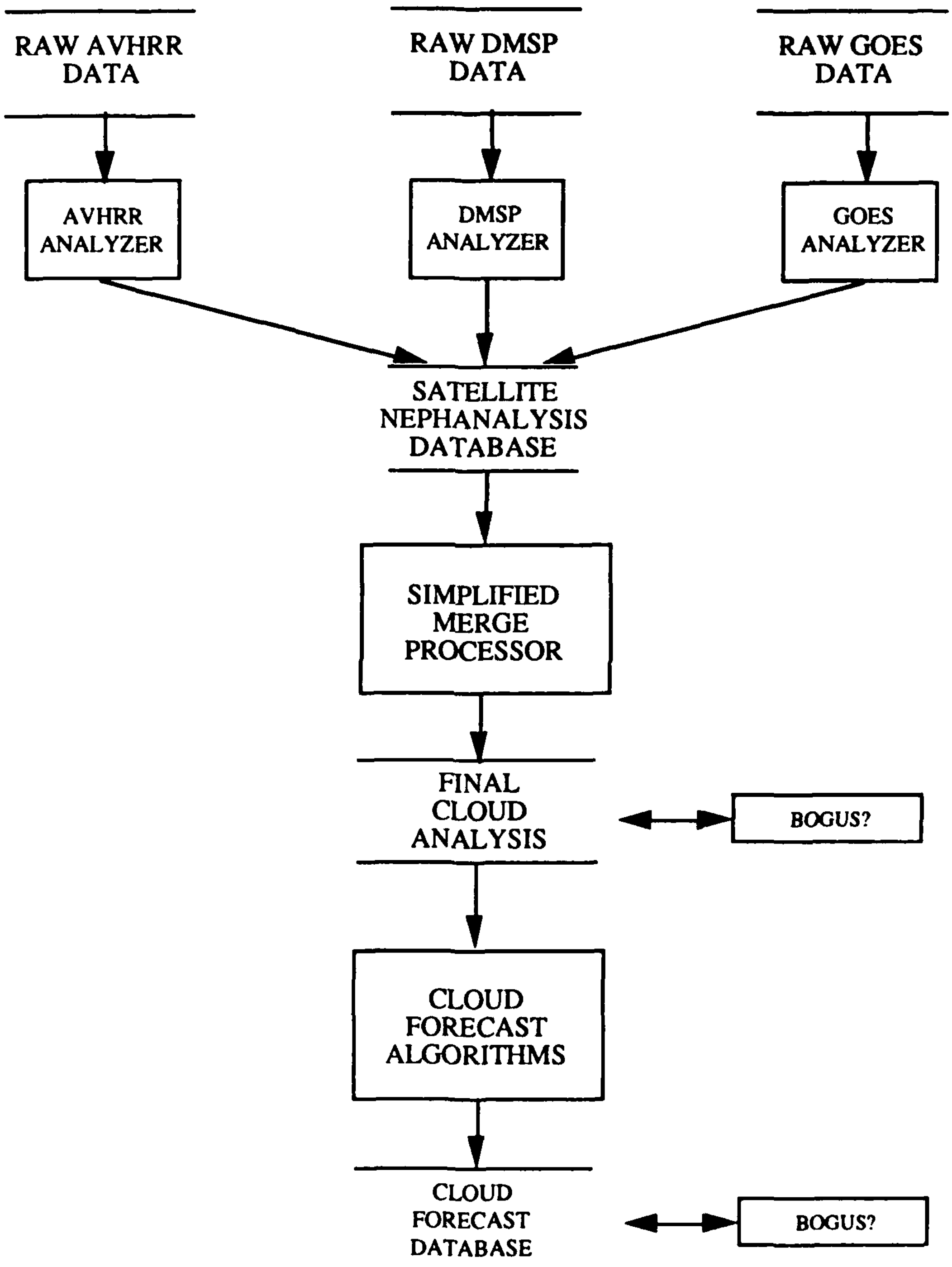


Figure 9: Illustration of the data and processing flow for the "extension of current RTNEPH design" algorithm.

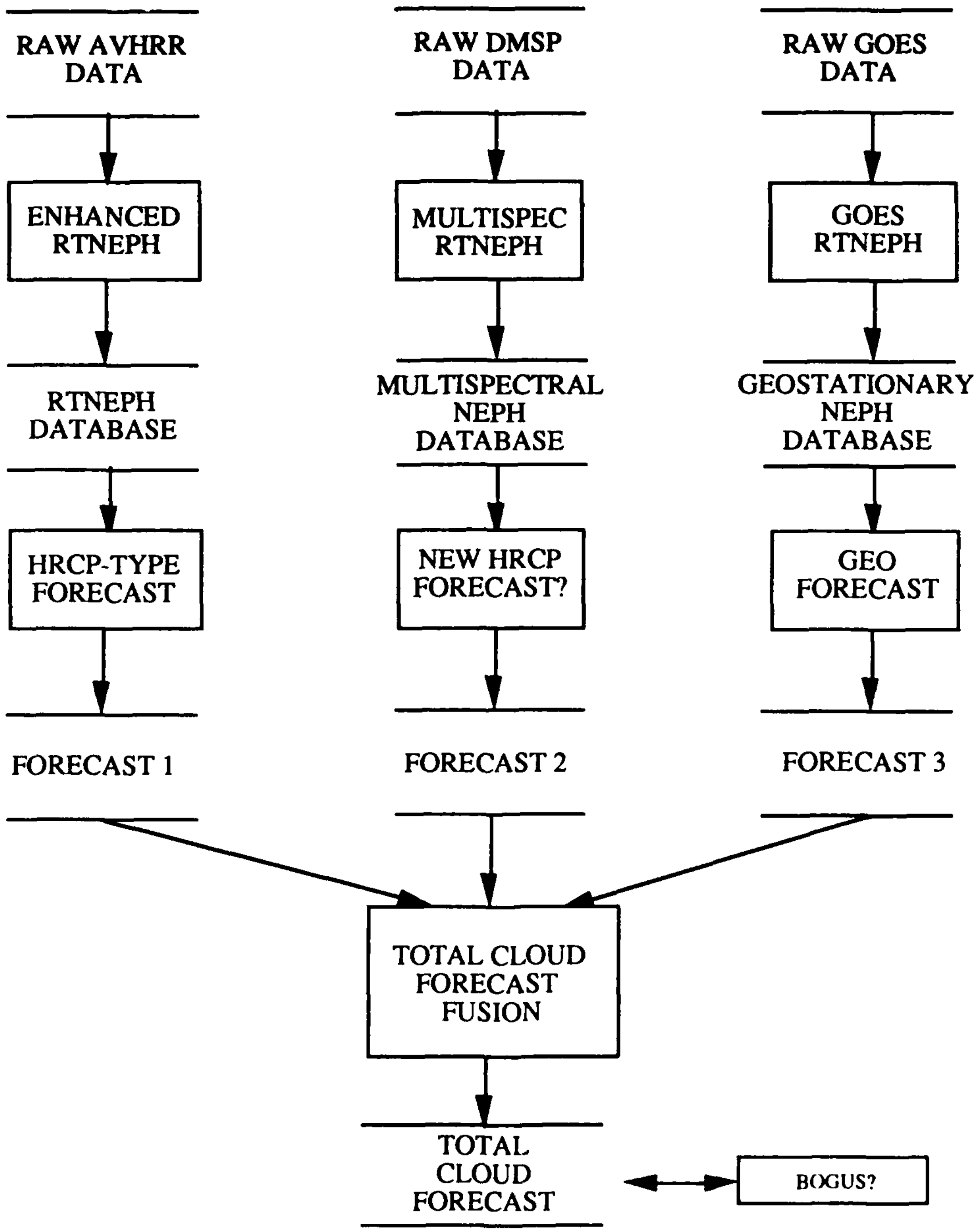


Figure 10: Illustration of the data and processing flow for the "forecast integration" algorithm.

(3) Disadvantages: Synthesis of a total cloud forecast from many sources would be relatively easy, but synthesis of a multi-parameter forecast is more difficult. Further, the users desiring a synoptic, synthesized MSNEPH database would be neglected.

### *5.1.5 Proposed Optimum Design*

We believe that algorithm 2, with some minor modifications, allows the most flexibility, the most accuracy, and though complex, still has a reasonable chance of being fully prototyped despite the short timelines. Possible enhancements to algorithm 2 would be the archival of the individual DMSP, GOES, and AVHRR analyses for climate users, and the implicit addition of a cross-correlations forecast ability in the suite of cloud forecast algorithms. The development would proceed from a simple scheme to a more robust scheme so that a baseline algorithm could be delivered even if the full prototype could not.

## *5.2 Cloud Analysis Integration Agorithm: Design Approach*

We now sketch one possible design for the cloud analysis integration algorithm. We assume that the individual nephanalyses for each satellite are stored in a database available to the integration algorithm. For these individual nephanalyses, no information is implied, e.g., no overlap assumptions are applied, nor are non-analyzed parameters such as cloud base filled in. An estimate of accuracy will be supplied along with each analyzed value. The purpose of the integration algorithm is then to optimally synthesize all the various satellite analyses into one complete nephanalysis. In the folowing sections, we will sketch a rough outline of the components of the integration algorithm and provide some background on the underlying methodology. It is likely that the algorithm will blend a knowledge-based approach with a technology similar to the optimum interpolation technology (Gandin, 1963; Schlatter, 1975; Lorenc, 1981) used in numerical weather analysis. Optimum interpolation (OI), to be described in more depth later, is a commonly used but computationally expensive technique for synthesizing observations into an analysis based on their accuracy.

### 5.2.1 *Use of Knowledge-Based Approach*

For computational reasons, it will be appropriate to apply the OI only when necessary. Thus, a set of rules will be necessary to determine when to use the OI, how to set up the data as input, how to define values for parameters not amenable to OI (such as cloud type), and how to consistency-check the analysis. These rules could be phrased in the same way as in the existing RTNEPH merge processor, or the IF-THEN decisions might be phrased as a set of rules using artificial intelligence inference engine technology. As a simple example, consider the analysis of total cloud. Rules might guide the use of OI here, such as:

(1) In the absence of new data the old total cloud should simply be persisted, or filtered through a forecast algorithm rather than the OI.

(2) If only one timely satellite analysis exists, or if all analyses agree on cloudy or clear conditions, then no optimum blending is required.

(3) Assume two competing analyses are input to the cloud analysis integration algorithm, a recent, highly accurate AVHRR analysis and an older, less accurate DMSP analysis. There is no reason to include any influence of the DMSP data despite its existence; use the most recent data alone.

Let's assume the existing RTNEPH methodology of IF-THEN statements is to be used. In this case, the flowchart in Figure 11 shows how a high-level design combines the rules above and the OI, performing the OI only if certain conditions had been met.

Were the three rules above the only sensible ones needed, then simple IF-THEN constructs would be the obvious technique to use. Unfortunately, there is more than just total cloud to analyze; the layer structure must be determined and consistency-checked. Even the current, rather simple merge algorithms of the existing RTNEPH use hundreds of IF-THEN constructs strewn throughout many subroutines to accomplish this. This will undoubtedly become even more complex with the greater variety of data sources to be synthesized in the SERCAA project.

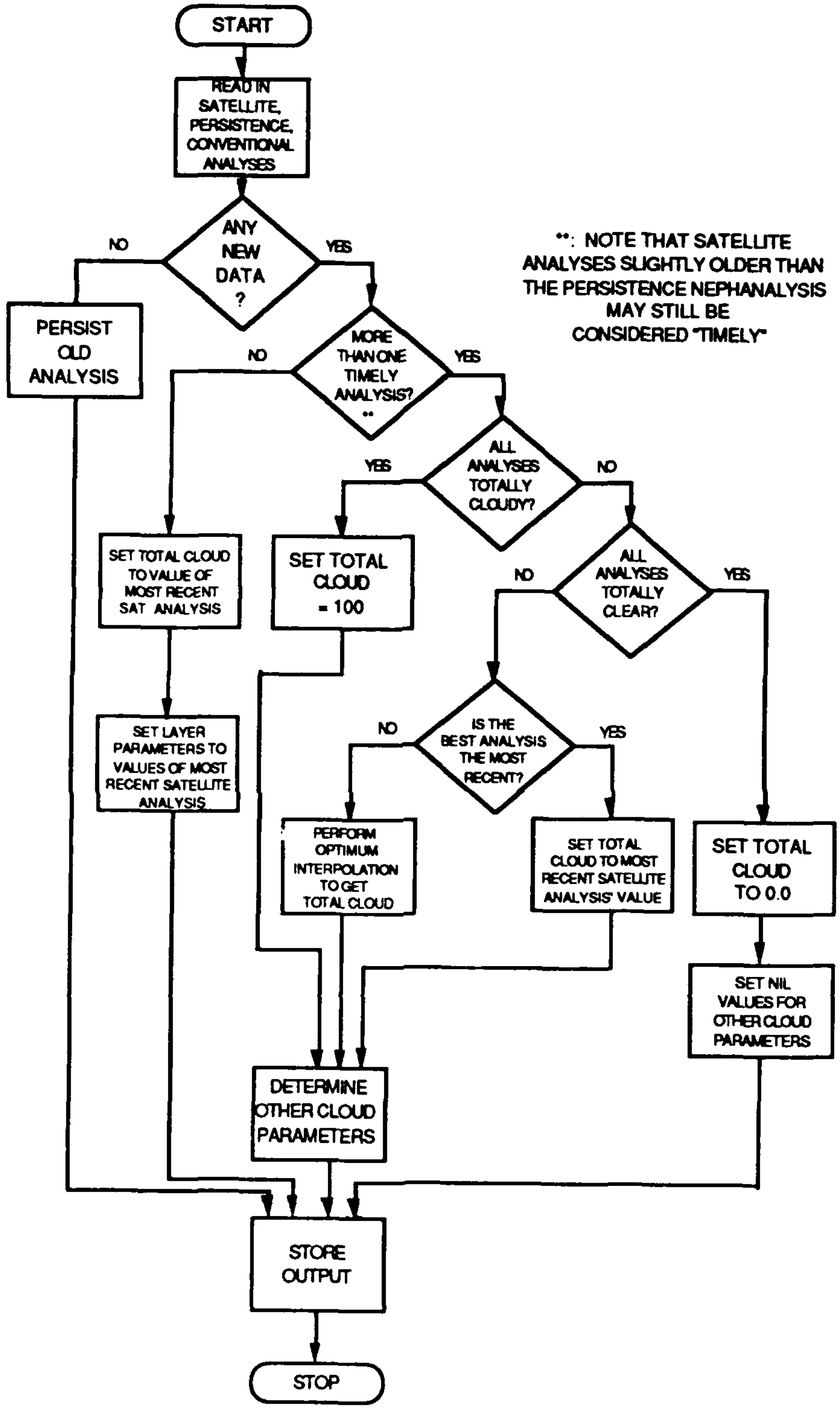


Figure 11: High-level flowchart illustrating the nesting of optimum interpolation algorithms for determining total cloud amount into a framework of IF-THEN decisions.

Modern knowledge based artificial intelligence (KBAI) provides a potential answer to this problem. KBAI is now usually implemented using a so-called "inference engine," which applies the propositional (predicate) calculus (formalized rules of logic) to arrive at a goal (e.g., the answer to the question: which data should be passed along to the OI) using facts (from the cloud data base) and knowledge codified as IF-THEN propositions. The inference engine is a standardized, commercially available computer program. The knowledge base is essentially a list or data base of propositions such as the three rules listed earlier. If some newer, more complete propositions are determined, there is no new programming; one simply edits the knowledge base and adds the new rule(s). Standard KBAI systems now come complete with natural language interfaces and procedures for diagnosing how decisions were made. Procedures for developing the knowledge base are not yet completely automated. These systems are most popular for providing guidance in narrow fields of expertise. They are basically clinicians. Such a system should serve well in quality controlling the large array of available cloud data. However, the interface between the cloud data base and the KBAI may require some custom work. Another novel aspect of the current project is the large number of mesh points. If possible, the system should track the decisions made at neighboring mesh points since cloud formations often spread out horizontally for large distances. For example, a large clear area would not need to be checked point-by-point, but rather considered as a group. This "connectivity" requirement may be applied in other ways and will be discussed in more depth later.

Many complex rules will be necessary to define layer parameters. Cloud layers must be sorted by height and have internally consistent types, bases, tops, and amounts. For example, stratus must be assigned a low altitude, and a cumulus cloud type must not be given 100 percent coverage. The algorithmic details for determining layer parameters will largely depend on the customer's desire for a pure satellite-sensed database, or a database with conventional data and layer structure not directly measured by the satellite included, such as cloud overlap assumptions. If only satellite-sensed information is desired, the approach could be relatively simple; for example, the algorithm would work downward layer-by-layer, determining an amount and type (probably using rules) and then setting top, base, and thicknesses layer-by-layer through a multivariate OI, described later. The summed layer cloud amounts would be subject to a consistency check with the total cloud, and the layer tops and bases checked to ensure no overlap in the vertical.

However, if a robust, three-dimensional cloud analysis is desired, with accurate inferences made on the structure of hidden layers, the technique must be more complex. In such a case, conventional data would likely be used to augment the analysis of low cloud and cloud base, and some of the rules for a satellite-only analysis would be discarded, such as the layer amounts summing to the total cloud amount. The combination of a knowledge-based system and OI methodology would probably be preserved, sorting out the simple cases from those requiring use of OI.

Using conventional data more carefully than is done with the current RTNEPH may dramatically improve the overall nephanalysis. Currently the RTNEPH uses all current conventional observations possible and applies a spreading algorithm to increase the number of points benefited by the conventional data. Some more judicious use of the conventional data might dramatically improve low cloud definition. First, because of their unreliability, nighttime conventional cloud observations might be thrown out (perhaps excluding those reporting rain, snow, or fog, which can be accurately sensed at night). Observations should be checked for consistency with the satellite data; if radically different than the satellite data for a cloud type that is typically well-analyzed by satellite, the conventional observation would be discarded. Further QC rules may be applied.

Remaining conventional observations would likely have much more reliable information on cloud base than could be inferred from satellite. Thus, estimates of low cloud amount and the cloud base from conventional data could be used more liberally than in the current RTNEPH. For example, consider a point with a low cloud layer and its base accurately measured by a ground-based observer and an opaque high cloud deck observed by satellite. Many other surrounding points with similar high clouds and similar underlying terrain would likely have a very similar low cloud amount and cloud base. This methodology could become more complex for an area like Europe, where conventional data is plentiful; in such a case, conventional observations may need to be objectively analyzed so that a conventional cloud analysis is available gridpoint-by-gridpoint, colocated with the satellite observations.

### *5.2.2 Use of Optimum Interpolation Methodology*

As mentioned previously, OI is a numerical analysis technique for synthesizing multiple observations into a single, consistent, accurate analysis. Accuracy is achieved by a judicious weighting of the data based upon the error characteristics of the observations; error-prone observations are deweighted, and highly reliable observations heavily weighted. We assume the accuracy will be calculated or estimated in the previous satellite sensor analysis processing. The OI technique was originally designed to optimally determine values in varying fields to initialize numerical forecast models. Input data for such an OI typically includes radiosonde data and surface and satellite observations, and the output data would be a smooth, continuous field of temperatures, geopotential heights, etc. For cloud analyses, smoothness is not desirable, and cloud observations to be combined will be scattered in time but not in space. This allows some simplifications to be made to the optimum interpolation methodology, which typically must account for covariance of errors between closely spaced observations of the same type. In our case, we will use only one observation of each type; for example, only the AVHRR analysis at point (I,J) will determine the cloud analysis at (I,J). The AVHRR analysis at (I+1,J) will not be included, i.e., satellite analyses at nearby gridpoints will have no influence. We will also make the simplifying assumption that there is no correlation of the error characteristics of the different cloud analyses.

#### 5.2.2.1 Univariate total cloud analysis example

We illustrate now the simplest possible example of optimum interpolation technology, the assimilation of two observations (or one observation and a first guess). This could be, for example, the determination of a new total cloud amount from the AVHRR and DMSP total cloud amounts.

The concept of analysis and data quality is key to this discussion. As shown below, in an optimal (weighted average) combination of two pieces of data, the ratio of the weights will be the inverse of the ratio of the square of the associated standard deviations. Thus if one datum has a standard deviation ten times as large as the other, its optimal weight will only be one hundredth of the other. Every analysis quantity, either input or output of the cloud analysis integration should have an associated accuracy or standard deviation. We use the term "accuracy", A, denoted in the special sense of

$$A = 1/v \tag{5.1}$$

where v is the variance (standard deviation squared). For an optimal combination of two independent measurements, the accuracies add. This is seen in the example which follows. Here we attempt to analyze cloud amount $C_a$ using cloud observations $C_1$ and $C_2$, which are assumed to be free of bias. They have prespecified accuracies $A_1$ and $A_2$. In such case, the analyzed cloud is determined by:

$$C_a = w\, C_1 + (1\text{-}w)\, C_2 \tag{5.2}$$

where w is the weight (0.0 to 1.0) applied to observation C1. We seek to determine the weight w. Now, assuming no covariance between observation types, we can write

$$\mathbf{V_a}' = w^2\mathbf{V_1}' + (1\text{-}w)^2\mathbf{V_2}' \tag{5.3}$$

where **V** indicates standard deviation squared as before. Now, minimizing this with respect to 2 yields

$$0 = 2w\mathbf{V_1}' + 2^*(1\text{-}w)\mathbf{V_2}' \tag{5.4}$$

or

$$w = \mathbf{V_2}' / (\mathbf{V_1}' + \mathbf{V_2}') \tag{5.5}$$

Since accuracy was predefined as the inverse of the variance, this can be rewritten as

$$w = A_1 / (A_1 + A_2) \tag{5.6}$$

This determines the relative weights. Now notice that we have also a quantitative expression for the analysis accuracy. Rewriting (5.2) and substituting (5.6) we have:

$$(A_2 + A_1)\, C_a = A_1\, C_1 + A_2\, C_2 \tag{5.7}$$

and it can be shown (Lorenc, 1986) that $(A_2 + A_1)$ is the accuracy of the final analysis, or $A_a$.

Now it is easy to extend the concept to three or more observations using the same technique. Consider Ca now to be an observation with accuracy Aa, and a new observation $C_3$ uncorrelated with anything previous. In this case, a new analysis of cloud amount, $C_n$, is determined by:

$$(A_3 + A_a)C_n = A_3C_3 + A_aC_a \tag{5.8}$$

5.2.2.2 Multivariate optimum interpolation of layer parameters

Determining an optimum total cloud is much simpler than determining an optimum layer structure. Lower cloud can be obscured by upper cloud, and layer type cannot be related easily to, say, layer thickness, allowing a multivariate OI treatment. Again the rules-based approach will be important for sorting out the times when OI can be applied and for resolving ambiguities, such as when to use the OI, what cloud type to use, how to determine the optimum number of layers, and so forth. For purposes of illustration, however, we show the potential benefit of a multivariate approach for determining the cloud layer top, base, and thickness together through an OI. OI provides a straightforward and optimal means of combining all three types of observations while maintaining the constraint that cloud thickness is the difference between cloud top and cloud base. When a linear constraint of this type is used in OI to derive covariances, then the analysis increments also satisfy the constraint. In the present case we specify the background covariances for cloud top and cloud base and use the constraint to determine all other background covariances. In general the covariances should be conditioned by the cloud type and possible geographical area.

Deriving the basic method, we follow Lorenc (1981) but do not scale the equations by the standard deviations. We let $B_i$ be any observed datum such as cloud base height, $A_k$ any analyzed value such as a cloud top or base height, $P_i$ and $P_k$ be the corresponding predicted (first guess) values, and $T_i$ and $T_k$ the corresponding true values. The true values do not include scales smaller than the analysis. We let

$$
\begin{aligned}
a &= A - T \\
b &= B - T \\
p &= P - T \\
q &= b - p = B - P \qquad &\text{(observational increment)} \\
r &= a - p = A - P \qquad &\text{(analysis increment)} \qquad (5.9)
\end{aligned}
$$

The method assumes

$$r_k = \Sigma w_{ik} q_i \qquad (5.10)$$

This is the same as

$$a_k = p_k + \sum_{i=1}^{n} w_{ik} (b_i - p_i) \qquad (5.11)$$

i.e., the analysis is a linear combination of weighted corrections between observations and the first guess. The squared expected analysis error is thus

$$
\begin{aligned}
\langle a_k2\rangle = \langle p_k2\rangle + 2\, \Sigma w_{ik}(\langle b_i p_k\rangle - \langle p_i b_k\rangle + \\
\Sigma\Sigma w_{ik} w_{jk} (\langle b_i b_j\rangle - \langle b_i p_j\rangle - \langle b_j p_i\rangle + \langle p_i p_j\rangle) \qquad (5.12)
\end{aligned}
$$

We now set the derivative of this with respect to $w_{ik}$ to zero, or simply note that

$$<a_k q_j> = 0.0, \tag{5.13}$$

i.e., the analysis error should not be correlated with the observations. Thus

$$0 = <a_k q_j> = <p_k q_j> + \Sigma w_{ik}<(b_i - p_i)q_j> \tag{5.14}$$

Therefore,

$$\Sigma w_{ik}(<b_i b_j>) - <b_i p_j> - <p_i b_j> + <p_i p_j>) = - <p_k b_j> + <p_k p_j> \tag{5.15}$$

Assuming

$$<bp> = 0 \tag{5.16}$$

i.e., that the first guess and observational errors are uncorrelated we have

$$\Sigma w_{ik}(<b_i b_j> + <p_i p_j>) = <p_k p_j> \quad j=1...n \tag{5.17}$$

Here the individual covariances could be determined through careful analysis over a large set of similar cases. In matrix terms this can be rewritten as,

$$\mathbf{M}\mathbf{w}_k = \mathbf{h}_k \tag{5.18}$$

or, solving for **w**,

$$\mathbf{w}_k = \mathbf{M}^{-1}\mathbf{h}_k \tag{5.19}$$

yielding finally,

$$\mathbf{r}_k = \mathbf{w}_k{}^T.\mathbf{q} \tag{5.20}$$

where k denotes the kth analyzed variable.

We now show an example of the multivariate OI of cloud top (T), base (B), and thickness, or depth, (D), where D = T-B, all performed at a single location. The first guess here is

$$T = 2500 \qquad \langle tt \rangle = 400^{**}2$$
$$B = 2100 \qquad \langle dd \rangle = 500^{**}2 \tag{5.21}$$
$$D = 400$$

It is assumed in an a priori analysis we have determined that the errors in T and D are correlated, with r=-0.4. Thus,

$$\langle td \rangle = 400*500*(-0.4). \tag{5.22}$$

The other covariances are determined by applying the constraint

$$d = t\text{-}b, \tag{5.23}$$

i.e.,

$$\langle bb \rangle = \langle (t-d)(t-d) = \langle tt \rangle - 2\langle td \rangle + \langle dd \rangle \tag{5.24}$$

Assume we now have three observations and estimates of each observation's error:

$$\begin{array}{ll} (1)\ T = 3000, & \langle tt \rangle = (1000)^{**}2 \\ (2)\ B = 1800, & \langle bb \rangle = (100)^{**}2 \\ (3)\ T = 2800 & \langle tt \rangle = (250)^{**}2 \end{array} \tag{5.25}$$

These observation's errors are all assumed to be uncorrelated with each other. The observational increments expressed as a vector are thus

$$\mathbf{q} = (500, -300, 300)^T \tag{5.26}$$

The analysis vector is

$$\mathbf{r} = (r_t, r_d, r_b) \tag{5.27}$$

or top, depth, and base. The matrix **M** is

$$\mathbf{M} = \begin{pmatrix} 116 & 24 & 16 \\ 24 & 58 & 24 \\ 16 & 24 & 22.25 \end{pmatrix} * 100^2 \tag{5.28}$$

and the right-hand side vectors are

$$H = (h_t, h_d, h_b) = \begin{pmatrix} 16 & -8 & 24 \\ 24 & -33 & 57 \\ 16 & -8 & 24 \end{pmatrix} * 100^2 \qquad (5.29)$$

Now, solving **Mw** = **H**, the solution for the weights is

$$w = (w_t, w_d, w_b) = \begin{pmatrix} 0.029871 & 0.027834 & 0.002036 \\ 0.203665 & -0.76476 & 0.968431 \\ 0.477936 & 0.445349 & 0.032586 \end{pmatrix} \qquad (5.30)$$

Applying (5.20) the analysis increments are,

$$r = ( 97.21656, 376.9518, -279.7352)^T$$

which must (and do) satisfy the constraint $r_d = r_t - r_b$. The final answer satisfies the constraints and balances the different data and predictions.

### *5.3 Implementation Strategy*

We have outlined an ambitious approach to cloud analysis integration with many complex algorithms involved. With a deadline of supplying a workable algorithm by May 1994, it will be important to concentrate on developing a workable core algorithm and adding complexity as time permits. Roughly, we plan to order the work as follows:

(1) Define the software and hardware needed for development. We will need to find an artificial intelligence language that can work in concert with a more conventional language such as Fortran or C. We may also need use of a data visualization language such as IDL or PVWAVE.

(2) Define the satellite analysis structure and database parameters necessary for the cloud analysis integration. This will allow SERCAA satellite algorithm development to proceed. Our current thoughts are to have the satellite cloud analyses store a pixel-by-pixel map of cloud/no cloud in its

native projection, and associated with each pixel would be a pointer to a particular cloud type/height in a lookup table. However, the parameters eventually must be mapped to a regular grid. Before the analysis integration, there would be an intermediate step of locating the pixel nearest the center of each output gridpoint and then determining the analysis parameters by summing up over a range of pixels around the central pixel.

(3) Decide on evaluation methodology, i.e., how the technique is to be tested and verified. Should work start with synthetic data in lieu of complete nephanalyses, or should work wait until real data is available? Currently we are leaning toward developing a prototype algorithm designed to work on an individual analysis point at a time, and then test primarily with synthetic cloud analyses as input. There are numerous advantages to this approach. First, since suites of cloud analyses will not be ready until late in the SERCAA project, this will allow testing to begin without real data. Second, synthetic data can be made to vary over a wider range of conditions than actual satellite data, allowing for a more robust test of the algorithm. Last, by working point-by-point, as we plan, the coding complexity is minimized due to much simple I/O, and a reasonable prototype should be able to be delivered under the tight timelines of the SERCAA effort. What is originally missing from this approach is up-front visualization of the effects of the integration algorithm applied to a full scene. Time permitting, we hope to apply the point analysis integration algorithm to a full array of points so that the algorithm output can be visualized with actual satellite data as input. (It is understood that large scale real data testing will be required prior to acceptance by operational users.)

(4) Prototype a simple analysis integration module, e.g., one which reads in all data, determines the most current data, determines missing values, and overwrites old analysis with new data.

(5) Start working on more complex parts, beginning with simpler tasks and tasks that will affect large numbers of analysis points, and proceeding to complex tasks and those that will affect smaller numbers of gridpoints. For example, when we are ready to work on optimum interpolation, we will start

with the methodology for total cloud, then multivariate for layer data, and then for data overlap and conventional data treatment.

### *5.4 Advanced Techniques: Cloud Overlap Processing*

A number of interesting and potentially beneficial techniques could be added to the basic integration algorithm. For example, in the future it may be possible to enhance areas having degraded analyses because of low-quality satellite data through relationships between satellites. We call this "overlap processing." The overlap processor makes use of cloud connectivity to spread the influence of the best data sources. The key concept here is the ability of satellite imagery to delineate connected cloud pixels, thereby identifying horizontally extended cloud features or cloud masses. The knowledge that a given mesh box is within a given cloud mass is then used to condition further analysis. Basically all pixels within a cloud mass are much more likely to be similar. In the ordinary analysis problem distance is measured in the usual geometric sense. In the cloud analysis problem, all pixels in a given cloud mass are close or nearby to each other. Further sensor information can be processed more exactly when conditioned by this knowledge. When the same cloud mass is sensed by different overlapping instruments, the overlapping processor will be able to extend the enhanced analysis of the overlapped region to the rest of the cloud mass. The same processing approach will also allow maximal spreading of the information content of surface and aircraft observations and perhaps experimental nadir looking satellite instruments (stereo imagers, incoherent lidars, etc.).

Given an existing nephanalysis, one implementation approach to overlap processing is to first run the nominal system, producing the direct nephanalysis. In overlapping regions, the poorer quality will not have been used. That is, we base our approach on comparing poor quality withheld data to the current direct analysis, in regions where this analysis is very good. Typically, these will be regions where recent data from a superior instrument was obtained. Since as the analysis ages its quality decays, this approach will naturally handle non-simultaneity of data sources. Only poor quality data will be used in the overlap processor to avoid troublesome observation error correlations in the overlapped nephanalysis. Further, it is the poor quality data which we can enhance the most and poor quality data would not have improved the direct nephanalysis significantly.

A first critical step in the overlap processing is to segment the imagery into cloud masses. In a simple segmentation, one can simply grow clouds by attaching all connected pixels with similar cloud characteristics. Of course some cloud formations have significant horizontal extent without all cloudy pixels physically touching. For example, a field of trade cumulus, or a cirrus field exhibiting waves are not continuous cloud fields, but are continuous cloud masses for our purposes. Automatic scene segmentation is a requirement for the overlap processor.

Assuming now that suppose we have identified a cloud mass from relatively poor quality imagery. The next step is to assemble all the high quality analysis data within that cloud mass. Only very high quality analysis data will be used in the enhancement. Generally, these data will correspond to the recent overpass of a better quality imager. For certain properties (e.g. cloud base) derived primarily from surface or aircraft observations, perhaps only a handful of mesh points will have high quality data.

The next step is a perfect prog technique, trying to predict the high quality analysis data from the colocated SDR and raw EDRs of the lower quality data source. That is, we will perform a regression analysis of the high quality direct nephanalysis data in terms of the imagery and derived quantities of the lower quality data source. This regression analysis will be confined only to high quality "observations" and to a single cloud mass. The regression relationship will then be used outside of the overlap region, but within the same cloud mass to predict the analyzed quantities. These predictions are the overlapped EDRs, which will be used by the second or overlapped nephanalysis.

Two things to note: First a constant prediction may be very useful. For example, cloud base is at 2000 m. Second, the residuals of the regression should be examined for geographic trends. If there is a significant trend with position the overlapping relationship is untrustworthy. A relationship including the trend might be developed, but it would be too dangerous to extrapolate. Anyway, the estimated errors of the extrapolated relationship would be very large, making the extrapolated relationship useless.

## 6. Conclusions

This technical report has outlined some of the crucial issues and technologies relating to the cloud analysis integration in the MSNEPH designed under the SERCAA project. The MSNEPH will continue to be used primarily for initialization of AFGWC's short-range cloud forecast models, but at the same time should be designed with the future needs of the climate community in mind. With the design of AFGWC's future cloud forecast models unspecified, the particulars of many of the MSNEPH integration algorithm details are difficult to specify. This technical report reviewed the existing forecast models, some likely technological enhancements to the forecast models, and the impacts on the MSNEPH design. Due to the current lack of specific guidance, the optimal approach at this time is to develop an MSNEPH algorithm and database structure that is flexible and useful to the broadest range of forecast applications possible.

## References


Air Force Global Weather Central, 1992: DRAFT, Operational Requirements Document (ORD) for cloud depiction and forecast system (CDFS) II, USAF ORD 005-92-1.

Coakley, J. A., and F. Bretherton, 1982: Cloud cover from high resolution scanner data: Detecting and allowing for partially filled fields of view. *J. Geophys. Res*, **87**, 4917-4932.

Crum, T. D., 1987: AFGWC cloud forecast models. AFGWC Tech Note 87/001, Air Force Global Weather Central, Offutt AFB, NE, 73 pp.

Ek, M., and L. Mahrt, 1989: A User's Guide to OSU1DPBL Version 1.0.3, A One-Dimensional Planetary Boundary Layer Model (Draft). Dept. of Atmospheric Sci., Oregon State University, 190 pp.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Field,* Gidrometeorologicheskoe Izdatel'stvo, Leningrad, Translated from Russian, Israel Program for scientific translation, 1965, 242 pp.

Hamill, T. M., and T. Nehrkorn, 1993: A short-term cloud forecast scheme using cross correlations. Preprints, 9th AMS Conference on Interactive Information and Processing Systems, Anaheim CA, 96-101.

Hamill, T. M., R. P. d'Entremont, and J. T. Bunting, 1992: A description of the air force real-time nephanalysis model. *Wea. Forecasting*, **7**, 288-306.

Kiess, R.B., Callahan, K.P., Lanicci, J.M., 1993: The Air Force Global Weather Central's Short Range Cloud Forecast Model: Preprints of the 5th Conference of Aviation Weather Systems, AMS.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, ***112***, 1177-1194.

Mitchell, K. E., and D. Hahn, 1989: Development of a cloud forecast scheme for the GL baseline global spectral model. GL-TR-89-0343, Phillips Laboratory/Geophysics Directorate, Hanscom AFB, MA. 147 pp. ADA231595

Rossow, W. B., E. Kinsella, A. Wolf, and L. Gardner, 1985: International satellite cloud climatology project (ISCCP) - description of reduced-resolution radiance data. WMO/TD No. 58, World Meteorological Organization, 132 pp.

Schiffer, R. A., and Rossow, W. B., 1985: ISCCP Global radiance dataset: a new resource for climate research. *Bull. Amer. Meteor. Soc.*, **66**, 1498-1505.

Schlatter, T. W. 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **104**, 246-257.

Slingo, J.M., 1987: The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, **113**, 899-927.

Snow, W., 1992: SERCAA Program Plan (30 October 1992). Phillips Laboratory/GPAS, 29 Randolph Rd, Hanscom AFB, MA 01731-3010. 52 pp.

Sundqvist, H. E. Berge, and J. Kristjanson, 1989: Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Mon. Wea. Rev.* **117**, 1641-1657.

Trapnell, R. N., 1992: Cloud-curve algorithm test program. PL-TR-92-2052, Phillips Laboratory, 29 Randolph Rd, Hanscom AFB, MA 01731-3010. 152 pp. ADA253918

Zivkovic, M. and J.-F. Louis, 1992: A new method for developing cloud specification schemes in general circulation models. *Mon. Wea. Rev.*, **120**, 2928-2941.